AD_____

Award Number: W81XWH-09-2-0175

TITLE: Proposal for Development of EBM-CDSS (Evidence-Based Clinical Decision Support System) To Aid Prognostication in Terminally Ill Patients

PRINCIPAL INVESTIGATOR: Benjamin Djulbegovic, M.D., Ph.D.

CONTRACTING ORGANIZATION: University of South Florida
Tampa, FL 33620

REPORT DATE: October 2013

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE
October 2013

2. REPORT TYPE
Annual

3. DATES COVERED
25 September 2012 – 24 September 2013

4. TITLE AND SUBTITLE
Proposal for Development of EBM-CDSS (Evidence-Based Clinical Decision Support System) To Aid Prognostication in Terminally Ill Patients

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-09-2-0175

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Dr. Benjamin Djulbegovic

E-Mail: bdjulbeg@health.usf.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of South Florida
Tampa, FL 33620

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

[illegible]

15. SUBJECT TERMS
CDSS, SUPPORT, DEALE, Terminally ill, Hospice, Prognostication

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	52	19b. TELEPHONE NUMBER *(include area code)*

Table of contents	
	Page number

Introduction

The goal of this project is to develop an Evidence-based Clinical Decision Support System (CDSS-EBM) available at the point of care which will improve prognostication of life expectancy of terminally ill patients and facilitate the hospice referral process. In addition, the CDSS-EBM will be expanded with an evidence based pain management module (EB-PMM) to assist physicians managing patients with pain.

Body:

Key research-related accomplishments (since the submission of previous annual progress report):

Currently, the study is being conducted at the Moffitt Cancer Center (MCC) and Tampa General Hospital (TGH).

[We submitted the required documents including the research protocol and informed consent forms to the scientific review committee at Moffitt Cancer Center (MCC) and secured the approval from this committee to open our study at MCC.

We have revised our study protocol and related study documents such as informed consent forms to reflect this change.

We submitted an amendment request to reflect this change in study sites to the University of South Florida's (USF) institutional review board (IRB) and have obtained authorization from USF IRB office.]

Our progress regarding the task outlined in the statement of work is as follows:

Task 5: Implementation of EBM-CDSS to calculate life expectancy and referral decision thresholds using decision curve analysis (DCA) and acceptable regret (ARg) models

- We completed training and submitted the required documents for our research personnel to complete MCCs' (and TGH) credentialing procedures. This step was essential to initiate the prospective phase of our study.
- We revised our case report forms and the EBM-CDSS software including its graphic user interface.
- Based on the feedback we obtained from our research personnel, PI experience and the feedback we have obtained from the referring physicians, we have revised the software and the user guide. [After at least 2 iterations we have finalized EBM-CDSS software and user guide.] This helped further improve standardization of our protocol to facilitate easier use of the EBM-CDSS and the user guide by our research personnel. (NB there is

continuing quality monitoring; depending on the feedback from the co-investigators, referring physicians and research staff "in the field", we will continue to revise and adjusted our software while retaining the fidelity of the study goals). The system also continues to be informed by additional theoretical knowledge, which we continue to further develop (see Appendix for the latest publications from this application).

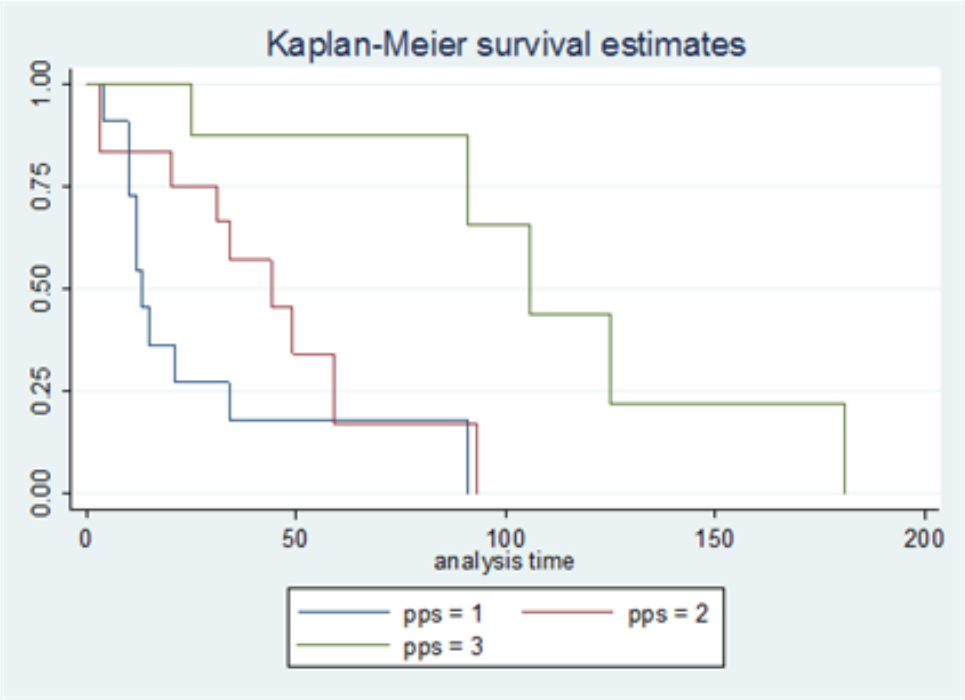
- We invested significant amount of time in training the research associates in using the EBM-CDSS software and fine tuning their interviewing skills. We conducted a number of mock interview sessions in which our research associates conducted interviews using the EBM-CDSS software, accompanying data collection forms, scripts and informed consents. [We had hired a research associate at our MCC site at the beginning of the year but after working with our team she left for pursuing further educational opportunities. Hence, we hired a new research associate to enroll patients and collect data at our MCC site. We made sure to have an overlapping period between the outgoing and incoming research associates.] The new research coordinator began on June, 2013. After completing the Collaborative Institutional Training Initiative's Human Subjects Research Curriculum, she was trained by the previous study coordinator for three weeks. Training included review of the study protocol, education on the existing literature regarding hospice, practice of the interview script, learning how to use the EBM-CDSS software for completing the interviews, as well as shadowing the previous coordinator. The new coordinator observed how to approach physicians for eligible referrals, obtain informed consent, read the patients' charts for lab work, conduct the interviews on the software, input the informed consents into Power Chart (a clinical data management software tool used at MCC), and make patients' folders so as to keep track of study participation and follow-up interviews. She also observed how to complete over the phone follow-up interviews. After practicing mock interviews with the previous coordinator, the new coordinator practiced with the study team. The previous research coordinator observed the new coordinator conduct her first interview with a patient. Following the interview, the previous coordinator shared some constructive criticism.
- We have used various strategies to raise awareness of our study to the referring physicians from the various specialties at MCC and TGH in order to improve enrollment of the patients in the prospective phase of our study. Specifically, as of June 2013, the hematology in-service, the hematology outpatient clinic, and the gastrointestinal clinic were the only clinics at Moffitt where physicians were referring patients to the study. After a couple of weeks of observing the patient flow and healthcare providers, the current research coordinator began to modify the study recruitment process. As opposed to only approaching the attending physicians before rounds regarding eligible patients, PI and research assistant asked to permission to join the rounds with the in-

service teams. PI introduces a research associate to the team and explained the study to all the nurse practitioners, physician assistants, interns, residents, fellows, and social workers on the service at different departments, at least on monthly basis. The similar process is followed at TGH, where the research team mostly works with palliative care service. Through this process, the number of patients potentially eligible and eventually enrolled in the study significantly increased. By establishing a relationship with the providers at TGH and MCC more patients have been referred.

- The PI and research coordinators have given a number of presentations to the referring physicians, social workers, nurses and staff at Moffitt and TGH to educate them on the study as well as ways that they can help with the referral process. This helped the awareness with the study in the Thoracic Clinic, Head and Neck Clinic and the Senior Adult Oncology Program at MCC as well as Palliative Care at TGH.
- We continue to regularly conduct meetings with TGH palliative care team and present the ongoing experience of our research study to the TGH palliative care team. These meeting established a fruitful and trustful and working relationship with TGH palliative care team, which is a key to facilitate the patients' referral to our study.
- As a result of these efforts we have enrolled 51 patients in our study. Specifically, at the TGH site we have screened 311 participants for eligibility, found 230 patients to be non-eligible for inclusion. We have enrolled a total of 32 patients at our TGH site. At our MCC site; we have approached 55 potential participants. Out of these 55 potential participants 24 participants were found to be ineligible. Out of the remaining 31 patients 19 patients have enrolled in the study.
- We have conducted an interim analysis based on the data collected on 31 patients. We evaluated the performance of PPS and SUPPORT prognostication models at 2 months. Calibration and discrimination statistic (the Brier score, scaled Brier score, the area under the receiver operating characteristic curve (AUC), and the Hosmer-Lemshow goodness-of-fit p-value) indicate that both PPS and Support performed well at predicting patient survival at day 60 (see table below). This provides the optimistic interim results that we are on right track and that indeed we will be able to develop the system which will facilitate better management in the end of life setting.

Statistic	PPS	PPS Modified	Support
Hosmer-Lemeshow P-value	0.39	0.31	0.35
Brier Score, Brier Score Scaled	0.19, 0.2	0.15, 0.3	0.23, 0.062
AUC (95% CI)	0.74 (0.54,0.94)	0.79 (0.58-1)	0.94 (0.82-1)

- A representative results (PPS score). For our 31 patients we grouped PPS (1 equals PPS = 30-40 for 11 patients, 2 equals PPS= 50-60 for 12 patients, 3 equals PPS = 70-90 for 8 patients). The results are rather encouraging: as the figure below shows, there is fairly distinct and stratified KM curves (log-rank test P = 0.003) between the subgroup of patients with different PPS score.



- We have refined our Evidence-based Chronic Pain Management Module to complement the CDSS-EBM. Our objective is to develop a reliable dosage conversion system as well as a knowledge base for each available pain medication. We have also incorporated evidence profiles for each drug to support the decision making using our pain management module. We have also created a survey to test usefulness of EB-PMM its users. The system is currently going through the final programming phase and it will be

first tested internally and then in the clinic in the prospective phase of the study. We have also created the user's manual for the EB-PMM.

- We developed an iOS (Ipad) based version of our EBM-PMM designed to assist physicians manage pain in adult cancer patients. The application includes the following functionalities:
    - Pain screening with standardized pain rating scale used to determine the patient's level of pain;
    - Selection of the appropriate medication based on to the levels of pain, type of patient (opioid naïve or opioid tolerant) and patient's preferences;
    - Calculation of total daily dose and single dose according to the medication presentation/concentration.
    - Conversion or rotation from one opioid to another opioid medication.
    - Prescription generation.

  We plan to test the usability and functionality of the application in our clinical sites.

- Drafted and submitted two manuscripts for peer-reviewed publication and published two manuscripts in peer-reviewed journals.

Reportable outcomes

1. Publications so far:
    - Eleazar Gil-Herrera, Ali Yalcin, Athanasios Tsalatsanis, Laura E. Barnes and Benjamin Djulbegovic, "Towards a Classification Model to Identify Hospice Candidates in Terminally Ill Patients", to appear in the Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2012
    - Miladinovic B, Kumar A, Mhaskar R, Kim S, Schonwetter R, et al. (2012) A Flexible Alternative to the Cox Proportional Hazards Model for Assessing the Prognostic Accuracy of Hospice Patient Survival. PLoS ONE 7(10): e47804. doi:10.1371/journal.pone.0047804
    - A. Tsalatsanis, I. Hozo, A. Vickers, B. Djulbegovic, "A regret theory approach to decision curve analysis: A novel method for eliciting decision makers' preferences and decision-making", BMC Medical Informatics and Decision Making 2010, 10:51 (16 September 2010)
    - A. Tsalatsanis, L. Barnes, I. Hozo, B. Djulbegovic, "Extensions to Regret-based Decision Curve Analysis: An Application to hospice referral for terminal patients", BMC Medical Informatics and Decision Making 2011, 11:77 (23 December 2011)

- E. Gil-Herrera, A. Yalcin, A. Tsalatsanis, L. Barnes, B. Djulbegovic, “Rough set theory based prognostication of life expectancy for terminally ill patients”, Proceedings of the IEEE EMBC 2011
- Mhaskar R, Miladinovic B, Tsalatsanis A, Mbah A, Kumar A, Kim S, Schonwetter R, Djulbegovic B. External Validation of Prognostic Models in Terminally Ill Patients. In: Hematology ASo, editor. American Society of Hematology Annual Conference; San Diego, California, 2011

2. Journal publications since last progress report: (appendix 1)
   - Jonathan M. Hernandez, Athanasios Tsalatsanis, Leigh Ann Humphries, Branko Miladinovic, Benjamin Djulbegovic, and Vic Velanovich, "Defining Optimum Treatment of Patients With Pancreatic Adenocarcinoma Using Regret-Based Decision Curve Analysis" to appear in Annals of Surgery, 2013
   - Wao H, Mhaskar R, Kumar A, Miladinovic B, Djulbegovic B. Survival of patients with non-small cell lung cancer without treatment: a systematic review and meta-analysis. Systematic reviews. 2013; 2(1): 10.
   - Miladinovic B, Mhaskar R, Kumar A, Kim S, Schonwetter R, Djulbegovic B. External validation of a web-based prognostic tool for predicting survival in patients in hospice care. Journal of Palliative Care, 2013.

Conclusion

We have already completed the majority of tasks described in the statement of work. We believe that we have closely followed the grant’s timeline where we could control the work process. At this point, we are focusing on enhancing enrollment of patients in our study and testing our Pain Decision Support System. To accomplish this, the PI will continue to carefully monitor the “situation on the ground” and further allocate distribution of the effort among the faculty and the staff from the available grant support to match the stated goals of our application.

Our key research findings so far can be summarized as follows:

- Based on results of our interim analysis we have confidence in the accuracy of predications of our prognostic models.
- However, these interim findings are based on small number of patients. We are working diligently to enroll more patients in our study. The efforts to enroll more patients will represent our key priority. The current strategies (see above) will be intensified, but we will likely need to add more research personnel to help with further increase in the

patient's accrual. This is necessary as this is time-intensive project and the current research personnel often cannot answer the referral requests as they are busy recruiting other patients.

- We are in the process of completing the Pain Decision Support System for Ipad platform. We are also in the process of internally testing the software for its accuracy, usability and acceptability by the end-users.
- Our goal is develop the appropriate theoretical framework that will facilitate the hospice referral process based on outcomes of multiple prognostication models. Our plan is to develop an evidence-based decision-support system for palliative, end-of-life care that will help both better referral to hospice as well help with pain management. Ultimately, the usefulness of our system will depend how well it performs when tested in clinical setting.

Next Steps

- Our immediate and most important next step is to enhance enrollment of patients in the prospective phase of the study. This requires tackling and coordinating multiple logistical, regulatory and administrative issues, which so far we have been successfully addressing. As explained above, we will likely need to hire new research personnel to help meet these goals.
- We will continue to work very closely with TGH palliative team and team of co-investigators from MCC to accomplish the goals of the study.
- We will maintain the quality assurance and oversight necessary for successful execution of the study.
- We will further develop and complete testing the EB-PMM (pain module). We will also pilot test our EB-PMM with physicians at Tampa General Hospital.
- Continue to contribute to knowledge base in this field by complete the on-going manuscripts and submitting them for publication in peer-reviewed journals.

Appendix 1 Peer-reviewed journal publications

Defining Optimum Treatment of Patients with Pancreatic Adenocarcinoma using Regret-based Decision Curve Analysis

Running Title: Regret Decision Analysis in Pancreatic Cancer

Jonathan M. Hernandez MD[1], Athanasios Tsalatsanis PhD[2,3], Leigh Ann Humphries[1], Branko Miladinovic PhD[2,3], Benjamin Djulbegovic MD Ph.D.[2,3,4], Vic Velanovich MD[1]*

[1]Department of Surgery, Division of General Surgery, University of South Florida Tampa, FL, USA

[2]Center for Evidence Based Medicine, University of South Florida, Tampa, FL, USA

[3]Department of Internal Medicine, Division of Evidence-based Medicine, Tampa, FL, USA

[4]Department of Hematology and Health Outcomes and Behavior, H. Lee Moffitt Cancer Center& Research Institute, Tampa, FL, USA

All authors declare that they have no conflicts of interest.

*Corresponding Author:
Vic Velanovich, MD
Department of General Surgery
1 Tampa General Circle, F145
Tampa, FL 33606
vvelanov@health.usf.edu

ABSTRACT

Background: Pancreatic adenocarcinoma is uniformly fatal without operative intervention. Resection can prolong survival in some patients; however, it is associated with significant morbidity and mortality. Regret theory serves as a novel framework linking both rationality and intuition to determine the optimal course for physicians facing difficult decisions related to treatment.

Methods: We used the Cox proportional hazards model to predict survival of patients with pancreatic adenocarcinoma and generated a decision model using regret-based decision curve analysis, which integrates both the patient's prognosis and the physician's preferences expressed in terms of regret associated with a certain action. A physician's treatment preferences are indicated by a threshold probability, which is the probability of death/survival at which the physician is uncertain whether or not to perform surgery. The analysis modeled three possible choices: perform surgery on all patients, never perform surgery, and act according to the prediction model.

Results: The records of 156 consecutive patients with pancreatic adenocarcinoma were retrospectively evaluated by a single surgeon at a tertiary referral center. Significant independent predictors of overall survival included preoperative stage ($p=0.005$, CI 1.19-2.27), vitality ($p<0.001$, CI 0.96-0.98), daily physical function ($p<0.001$, CI 0.97-0.99) and pathologic stage ($p<0.001$, CI 3.06-16.05). Compared with the "always aggressive" or "always passive" surgical treatment strategies, the survival model was associated with the least amount of regret for a wide range of threshold probabilities.

Conclusions: Regret-based decision curve analysis provides a novel perspective for making treatment-related decisions by incorporating the decision maker's preferences expressed as his/her estimates of benefits and harms associated with the treatment considered.

INTRODUCTION

Although significant progress has be made over the last two decades in reducing perioperative mortality for patients with localized pancreatic adenocarcinoma, pancreaticoduodenectomy remains associated with significant morbidity(1, 2). Moreover, long-term survival has remained unchanged and persistently elusive for the vast majority of patients with the disease(3, 4). Operative extirpation, for which about 15-20% of patients are eligible, is undertaken when technically feasible because it offers the only opportunity for prolonged survival, and because there are few alternative treatments – each of which has limited efficacy(5). However, even among patients undergoing complete tumor extirpation with negative margins, the disease recurs in 40% of the patients within 6 months, most commonly in the form of liver metastasis (6). These patients may derive little-to-no survival benefit from local control, while potentially suffering from operative morbidity(6). Selection of patients likely to benefit from aggressive local control is therefore particularly important in the management of patients with radiographic-localized pancreatic adenocarcinoma.

Decision analysis typically defines the probability of an event and provides the optimal model among alternative clinical management strategies, thus maximizing a definable outcome (7, 8). Probability models based on diagnostic and prognostic variables have been utilized to assist physician decision-making regarding various treatments and interventions, including resection for cancer, although the effectiveness of the models remains questionable(9-15). The reasons behind this skepticism include the probabilistic nature of these models that adds complexity to the decision process and, importantly, the reliance of most of these models on expected utility theory, which is often violated during decision making(16-20).

We recently developed a decision methodology that overcomes the limitations of probabilistic survival models, and which can be utilized to facilitate medical decisions based on the decision-maker preferences (19, 20). Our methodology, Regret-based Decision Curve Analysis or *Regret DCA*, relies on the cognitive emotion of regret to identify conditions under which a physician is unsure about the choice between alternative treatment strategies (19, 20). Surgeons, as with any decision maker, may experience regret (defined as the difference between the utility of an action taken and utility of an alternative action) if they eventually realize that a decision they made was suboptimal, and that an alternative form of treatment would have been preferable (21-27). Regret DCA utilizes this regret to compute the threshold probability at which the physician is uncertain about which treatment strategy to recommend to his/her patient. In this study, we used Regret DCA to facilitate treatment decisions for a cohort of patients with localized, resectable pancreatic adenocarcinoma.

The intention of this article is to present a novel decision methodology that relies on regret theory and attempts to explain medical decision-making for surgeons treating patients with pancreatic adenocarcinoma. Despite the fact that the prediction model presented has been well fitted to our data, its role in this article is secondary and its purpose is to demonstrate how the regret methodology can be used to evaluate three management strategies: aggressive, passive, or model-based decision making. In this context, we have demonstrated that the prediction model performs better the other two strategies in terms of regret.

MATERIALS AND METHODS

The records of 156 consecutive patients referred for surgical consultation from January 2005 to 2009 with pancreatic adenocarcinoma were retrospectively reviewed by a single surgeon at a tertiary referral center. The diagnosis was confirmed by histological evaluation, and disease stage was determined by pathological evaluation of the resected specimen and by imaging. All patients had been administered the SF-36 Health Survey to assess quality of life, which includes 36 statements grouped into 8 domains of quality of life: physical functioning, physical role, bodily pain, general health, vitality, social functioning, emotional role, and mental health. The SF-36 utilizes a Likert scale of 0 to 100, with higher scores indicating better/normal health or physical functioning. We previously

demonstrated that the SF-36 correlates well with pathology, survival, stage and resectability of pancreatic lesions (27).

The distribution for overall survival was estimated using the Kaplan-Meier Method. Cox proportional hazards modeling was used to determine the effect on survival of the following 12 covariates, including those described by SF-36: age, gender, stage, adjuvant therapy, physical functioning, role-physical, role-emotional, bodily pain, pretreatment vitality, mental health, social functioning and general health. Additional covariates such as tumor characteristics (lymphovascular invasion, perineural invasion, etc.) could potentially influence the output of the Cox model, however, this information is typically unknown to the surgeon *a priori*. Furthermore, such covariates were not included in the analysis since our dataset was originally constructed based on the methods and protocols designed for a study (28) focusing on the quality of life, pathology, resectability and survival in patients with pancreatic lesions. The model was created using stepwise elimination on all variables ($p< 0.15$ to enter, and $p< 0.20$ to stay). The proportional hazards assumption was examined using Schoenfeld residuals. The importance of each variable and the discriminative ability of the Cox model was examined using Royston-Sauerbrei's discrimination statistic D and explained variation R^2_D (29). All continuous variables were centered about the mean. All analyses were performed using STATA (30).

To derive the optimal treatment strategy, we then utilized the Regret-based Decision Curve Analysis methodology (*Regret DCA*)(19, 20). Regret DCA employs the decision maker's feeling of regret to compute the threshold probability at which he/she is uncertain about alternative actions, e.g., to operate or not to operate. In considering decisions for patients with pancreatic adenocarcinoma, we considered survival less than 7 months from the time of tumor extirpation as being unlikely to have imparted a survival advantage, and therefore unnecessary based upon median survival of patients with locally advanced, non-metastatic disease (31). Based on this assumption, we formulated a decision model that compares an individual patient's prognosis with the threshold probability at which the surgeon would be indifferent about recommending surgery.

Typically, decision theory suggests that a person should be treated if the probability of an event (i.e. the patient develops a disease; the patient dies; the patient survives longer than a predefined timeframe, etc.) is greater than or equal to a threshold probability (7, 8, 32). In this paper, we sought to treat the patients who were likely to survive longer than 7 months from the time of their resection. Therefore, the convention used is: **if the patient's probability of surviving 7 months is greater than or equal to the threshold probability ($s \geq P_t$), the surgeon should offer resection. If the patient's probability of survival is less than the threshold probability ($s < P_t$), the patient may be unlikely to benefit substantially from surgery and the surgeon should not recommend resection in favor of medical alternatives.**

The probability of survival can be computed for each patient based on the Cox survival model previously described. However, the threshold probability is subject to each surgeon's preferences and clinical practice attitudes. At the individual level, it can be computed as (19, 20):

$$P_t = \frac{1}{1+\frac{Regret\ of\ omission}{Regret\ of\ commission}} \qquad (1)$$

We define "*regret of omission*" as the regret felt by a surgeon who withheld necessary surgery from a patient who may have benefited from that resection (patients with localized disease who lived longer than 7 months). Conversely, "*regret of commission*" is the regret felt by a surgeon who performed an unnecessary surgery on a patient who derived no benefit from that operation (e.g. the patient died as a result of the procedure or died within 7 months from the time of resection). Both regret values can be determined using the Dual Visual Analogue Scales (DVAs) (Figure 1) (19, 20). Formally, regret can be expressed as the difference between the utility of the outcome of an action taken and the utility of the outcome of the action that, in retrospect, should have been taken (21-27). Commonly used

techniques for estimating utility, and therefore decision maker preferences, such as standard gamble and time trade-off are time consuming, cognitively complex and have been shown to lead to biased estimates of people's preferences (33-35). Instead, in this paper, we use the Dual Visual Analogue Scales (DVAs) to estimate directly the values of regret of commission and omission(19, 20). The DVAs comprise two 100-point scales, each anchored to no regret and maximum regret. One of the scales is used to elicit regret of omission and the other to elicit regret of commission (Figure 1).

After computing the surgeon's threshold probability, the clinical question regarding treatment for patients with pancreatic adenocarcinoma can be broken down into three strategies: 1. surgeons can stay passive and allow the disease to run its course, 2. surgeons can be aggressive and recommend resection on all patients, or 3. surgeons can use prediction model for guidance. Any of these strategies may cause regret if the outcome is poor. Under the Regret DCA methodology, the optimal strategy is the one that will cause the least amount of regret if that strategy is proven suboptimal. Formally, regret can be expressed as the difference between the utility of the outcome of the action taken and the utility of the outcome of the action that, in retrospect, should have been taken (21-27). Considering the decision tree that describes this clinical problem (Figure 2), we can compute the expected regret associated with each of the three strategies as follows:

$$ERg[NoSurgery] = (1 - s) * \frac{P_t}{1-P_t} \quad (2)$$

$$ERg[Surgery] = s \quad (3)$$

$$ERg[Model] = \frac{\#FP}{n} * \frac{P_t}{1-P_t} + \frac{\#FN}{n} \quad (4)$$

The values of #FP and #FN correspond to the number of false positive and false negative results, respectively, as compared to the actual patient outcomes used for the development of the prediction model, and the number of patients in the dataset is *n*. We define true positive (TP), true negative (TN), false positive (FP), and false negative (FN) results as follows:

TP: the number of patients who will survive longer than 7 months and for whom the estimated probability of survival is greater than or equal to the threshold probability (i.e., the patients who should receive surgery).

TN: the number of patients who will die in 7 months and for whom the estimated probability of survival is less than the threshold probability (i.e., the patients who should NOT receive surgery).

FP: the number of patients who will die within 7 months and for whom the estimated probability of survival is greater than or equal to the threshold probability (i.e., the patients who received unnecessary surgery).

FN: the number of patients who will survive longer than 7 months and for whom the estimated probability of survival is less than the threshold probability (i.e., the number of patients who should have received surgery but did not).

As shown in equations 2 and 4, the expected regret associated with each strategy is a function of the physician's threshold probability. To identify the least regretful action, the Regret DCA methodology computes the expected regret for a range of threshold probabilities (0-100), and expected regret is then graphed against the threshold probability for each of the three actions. The action with the lowest value of expected regret corresponds to the most desired action, given a certain threshold probability.

RESULTS

Patient Characteristics

A total of 156 patients with histologically-confirmed primary pancreatic adenocarcinoma were included. The mean age was 65.9 ± 10 years, 83% were stage I or II, 54% were resected, 66% received chemotherapy, and the median survival was 18 months (95% CI 12-26) (mean survival was 15.7 ± 25 months). The SF-36 scores revealed that role-physical and pretreatment vitality had the

lowest scores, and mental health had the highest score (Table 1). The distribution of overall survival is presented in Figure 3.

Survival model

Of the 12 variables included in the dataset, three met the stepwise inclusion criteria and were used to construct the survival model: stage, pretreatment vitality, and role-physical (daily physical functioning). The explained variation of the fitted model was R^2_D =0.4 (95% CI: 0.27-0.52) and the proportional hazard assumption were not violated (P < 0.96). Table 2 presents the estimates of hazard ratio for the Cox prediction model.

Regret Decision Curve Analysis

We employed Regret DCA to evaluate the three management strategies: 1. Recommend against potentially curative surgery in favor chemotherapy or chemoradiotherapy; 2. be aggressive and recommend resection, 3. use the prediction model as a decision aid. Figure 4 depicts the expected regret as a function of threshold probability for each of the three management strategies. As shown, the least regretful strategy for threshold probabilities greater than 5% is to utilize the prediction model. For threshold probabilities between 80-87%, the regret curve associated with the prediction model is subject to noise (36) that we attribute to the error term of the Cox prediction model. We assume that the prediction model remains the least regretful strategy within the 80-87% range as well. Our results demonstrate that the survival model we describe has significant clinical value for the majority of decision makers.

Hypothetical Case Study

A 72 year-old female with diabetes and hypertension has been diagnosed with pancreatic adenocarcinoma after undergoing endoscopic retrograde cholangiopancreatography (ERCP) and common bile duct stenting for obstructive jaundice. She is currently without pain and is tolerating a regular diet. Her jaundice resolved after the placement of her biliary stent. Her CT scan demonstrates a localized mass in the head of the pancreas without involvement of the superior mesenteric vein, portal vein, superior mesenteric artery, or hepatic arteries. The patient is active and able to perform all activities of daily living. She expresses a strong desire to spend as much time as she can with her grandchildren.

We demonstrate the decision process assuming two types of hypothetical decision makers: One surgeon is extremely selective in offering resection to patients with pancreatic adenocarcinoma (Surgeon #1), and the second surgeon (Surgeon #2) generally offers resection to all patients with radiographically-resectable disease. The process, depicted in Figure 5, is initiated with the elicitation of the surgeon's preferences. Using the DVAS method (Figure 1) we estimate the threshold probability as a function of regret of omission and regret of commission (equation 1). Suppose that the answers to the questions shown in Figure 1 for the surgeons are as follows:

Surgeon #1: Regret of omission: 20; regret of commission: 90. Therefore, the threshold probability is equal to: 81.8% (equation 1).

Surgeon #2: Regret of omission: 90; regret of commission: 4. Therefore, the threshold probability is equal to: 4.2%.

Based on the results of Regret-DCA (Figure 4), the optimal and least regretful strategy for Surgeon#1 is to use the prognostication model we developed, described above. If the patient's estimated probability of survival is greater than or equal to 81.8% (the threshold for Surgeon #1) then the optimal strategy is to treat (perform the operation). If the probability of survival is less than 81.8%, then the optimal strategy is to offer alternative treatments (forego resection). Conversely, for Surgeon #2, whose threshold probability is equal to 4.2%, the optimal and least regretful strategy is to offer resection.

As mentioned earlier, the Regret-DCA methodology can also be used by the patients (19). For completeness, we present how this process could work. The patient would be asked questions similar to those depicted in Figure 1. We have previously shown that patient ratings of utility scores closely correlate with quality of life after pancreaticoduodenectomy; moreover, this patient-centered assessment many change over time as quality of life improves (37).

Regret of omission: On a scale of 0 to 100, where 0 = no regret and 100 = maximum regret you could feel, how would you rate your level of regret if you did not have an operation that could have extended your life?

Regret of commission: On a scale of 0 to 100, where 0 = no regret and 100 = maximum regret you could feel, how would you rate your level of regret if you had an operation that did not extend your life?

DISCUSSION

We describe the theory and application of regret decision curve analysis as it applies to surgeons and to decisions regarding operative intervention in patients with pancreatic adenocarcinoma. To the best of our knowledge, this is the first application of regret DCA to assist surgeons in decision-making for patients with pancreatic malignancies. Our approach promotes personalized patient care by incorporating decision-maker preferences from the perspective of regret by estimating a threshold probability for a decision maker. We believe the decision regarding resection for patients with pancreatic adenocarcinoma is particularly well suited for a regret-based approach given the generally fatal prognosis for this disease, regardless of the decision made.

Modern cognitive theories seek to balance risks and benefits in the decision-making process by taking into account both intuition and analytical processes (37). We believe that rational decision-making should take into account both the formal principles of rationality and human intuition. We have accomplished this using regret, a cognitive emotion, to serve as the link between intuition and analytical thinking (19, 20). Eliciting surgeons' preferences by using regret is likely to prove superior to using traditional utility theory because regret explicitly forces the surgeon to consider consequences of decisions. Our method relies on elicitation of a threshold probability, which must be calculated for every decision maker. In other words, our model forces surgeons to consider the possible outcomes of recommending pancreaticoduodenectomy rather than simply recommending resection for all tumors that appear resectable on radiographic imaging.

We argue that our approach contributes to the field of decision-making, but we acknowledge that it is not a panacea. We do, however, believe that our methodology is best suited for medical decision-making primarily associated with trade-offs between quality and quantity of life. Pancreatic adenocarcinoma meets this criterion: surgical resection may offer an additional year of survival, albeit with the potential for serious morbidity, particularly if the resection is undertaken at low-volume centers (38, 39). For the fortunate 15-20% of patients with radiographically-localized disease amenable to resection, the median survival ranges from 17 to 23 months (40). At high-volume institutions with extensive experience, the mortality rate is <3%–5%, but morbidity remains problematic, with early postoperative complication rates of ~30%-40% (6). Perioperative morbidity and mortality rates recorded in national databases, which include data from a broad spectrum of hospitals and surgeons' experiences, report significantly higher numbers of complications than high-volume tertiary referral centers (38). Applying our model of regret theory may indirectly motivate each surgeon to consider their own results with the procedure and to consider the support available within the institution where the procedure is planned when contemplating the best course of action for each patient, further personalizing care.

A significant proportion of patients undergoing resection develop early metastatic disease and have very limited survival, and thus derive no benefit from the operative intervention (i.e., there is no trade-off improvement in quality-of-life). This issue has been addressed with the use of refined definitions of borderline resectability and the use of neoadjuvant therapy (41). Specifically, this minimally effective chemotherapy, which offers virtually no hope of eradicating disease and little if any therapeutic efficacy, does provide a "window of observation", during which distant metastatic disease may appear and thus spare the patient unnecessary surgery. This approach may minimize regret and results in better overall survival for patients who ultimately undergoing resection (42), but it has not been widely adopted across the country or even across academic centers. Similarly, regret theory remains severely underutilized in the healthcare arena, despite considerable conceptual and empiric interest in its applicability, and in the strong influence of regret on physician decision-making (32, 43-45). The lack of incorporation of regret theory into healthcare delivery is particularly perplexing, especially considering that all medical decisions are accompanied by varying degrees of risk and uncertainty, and – therefore – potential regret. Moreover, recent work has suggested that physicians' behavior can often be explained by regret avoidance (46), which further substantiates the need to incorporate regret modeling into healthcare decisions.

As with any novel theoretical work, our application of regret theory to pancreatic adenocarcinoma has limitations. First, we applied the theory retrospectively with assigned cutoff survival values. We assumed maximal regret to be associated with operating on a patient who died within the first seven months following resection. Excluding death as a result of the procedure (perioperative death), which is always associated with regret, death within seven months may not necessarily be associated with regret. For example, a patient may have died of an unrelated stroke that could not have been foreseen prior to resection. Second, our approach has not yet been empirically tested and the prediction model has not been externally validated. Third, the methodology, as presented, is appropriate for point decision-making, and not necessarily for decisions that re-occur over time – as frequently happens in patient care. Finally, we assumed that there is a single decision-maker involved in the process where, in actual practice, a multidisciplinary team of healthcare providers is involved in treatment decisions.

In conclusion, we have described a novel approach to surgical decision-making using the cognitive emotion of regret, which seeks to personalize care. The goal of our work is to power a computerized decision support tool to assist physicians and patients in making better medical decisions. We envision the tool to be shared by both physician and patient during consultation, in which the physician elicits the patient's preferences towards alternative management strategies.

ACKNOWLEDGMENTS

This work is partially supported by the Department of Army grant #W81 XWH 09-2-0175. The authors appreciate the assistance of Jane Carver of the University of South Florida Clinical and Translational Science Institute in reviewing this manuscript.

REFERENCES

1. McPhee JT, Hill JS, Whalen GF, Zayaruzny M, Litwin DE, Sullivan ME, et al. Perioperative mortality for pancreatectomy: a national perspective. Annals of surgery. 2007;246(2):246.

2. Vollmer CM, Sanchez N, Gondek S, McAuliffe J, Kent TS, Christein JD, et al. A root-cause analysis of mortality following major pancreatectomy. Journal of Gastrointestinal Surgery. 2011:1-15.

3. Baxter NN, Whitson BA, Tuttle TM. Trends in the treatment and outcome of pancreatic cancer in the United States. Annals of surgical oncology. 2007;14(4):1320-6.

4. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T, et al. Cancer statistics, 2008. CA: a cancer journal for clinicians. 2008;58(2):71-96.
5. Cress RD, Yin D, Clarke L, Bold R, Holly EA. Survival among patients with adenocarcinoma of the pancreas: a population-based study (United States). Cancer Causes and Control. 2006;17(4):403-9.
6. Beger HG, Rau B, Gansauge F, Poch B, Link KH. Treatment of pancreatic cancer: challenge of the facts. World journal of surgery. 2003;27(10):1075-84.
7. Pauker SG, Kassirer JP. Therapeutic decision making: a cost-benefit analysis. The New England journal of medicine. 1975;293(5):229.
8. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. The New England journal of medicine. 1980;302(20):1109.
9. Abramson MA, Pandharipande P, Ruan D, Gold JS, Whang EE. Radical resection for T1b gallbladder cancer: a decision analysis. HPB. 2009;11(8):656-63.
10. Aloia TA, Fahy BN. A decision analysis model predicts the optimal treatment pathway for patients with colorectal cancer and resectable synchronous liver metastases. Clinical Colorectal Cancer. 2008;7(3):197-201.
11. Kattan MW, Cowen ME, Miles BJ. A decision analysis for treatment of clinically localized prostate cancer. Journal of general internal medicine. 1997;12(5):299-305.
12. Kulkarni GS, Finelli A, Fleshner NE, Jewett MAS, Lopushinsky SR, Alibhai SMH. Optimal management of high-risk T1G3 bladder cancer: a decision analysis. PLoS medicine. 2007;4(9):e284.
13. Lotan Y, Cadeddu JA, Lee JJ, Roehrborn CG, Lippman SM. Implications of the prostate cancer prevention trial: a decision analysis model of survival outcomes. Journal of clinical oncology. 2005;23(9):1911-20.
14. Steyerberg EW, Marshall PB, Jan Keizer H, Habbema JDF. Resection of small, residual retroperitoneal masses after chemotherapy for nonseminomatous testicular cancer. Cancer. 2000;85(6):1331-41.
15. Telford JJ, Saltzman JR, Kuntz KM, Syngal S. Impact of preoperative staging and chemoradiation versus postoperative chemoradiation on outcome in patients with rectal cancer: a decision analysis. Journal of the National Cancer Institute. 2004;96(3):191-201.
16. Baron J. Thinking and deciding: Cambridge University Press; 2000.
17. Bell DE, Raiffa H, Tversky A. Decision making: Descriptive, normative, and prescriptive interactions: Cambridge University Press; 1988.
18. Dawes RM, Kagan J. Rational choice in an uncertain world: Harcourt Brace Jovanovich New York; 1988.
19. Tsalatsanis A, Barnes LE, Hozo I, Djulbegovic B. Extensions to Regret-based Decision Curve Analysis: An application to hospice referral for terminal patients. BMC Med Inform Decis Mak. 2011 Dec 23;11(1):77. PubMed PMID: 22196308. Epub 2011/12/27. Eng.
20. Tsalatsanis A, Hozo I, Vickers A, Djulbegovic B. A regret theory approach to decision curve analysis: a novel method for eliciting decision makers' preferences and decision-making. BMC Med Inform Decis Mak. 2010;10:51. PubMed PMID: 20846413. Pubmed Central PMCID: 2954854. Epub 2010/09/18. eng.
21. Djulbegovic B, Hozo I. When Should Potentially False Research Findings Be Considered Acceptable? PLoS Med. 2007 02/27;4(2):e26.
22. Djulbegovic B, Hozo I, Schwartz A, McMasters KM. Acceptable regret in medical decision making. Med Hypotheses. 1999 Sep;53(3):253-9. PubMed PMID: 10580533. eng.
23. Hozo I, Djulbegovic B. When is diagnostic testing inappropriate or irrational? Acceptable regret approach. Med Dec Making. 2008;28(4):540-53.
24. Hozo I, Djulbegovic B. Will insistence on practicing medicine according to expected utility theory lead to an increase in diagnostic testing? Med Dec Making. 2009;29:320-2.
25. Bell DE. Regret in Decision Making under Uncertainty. Operations Research. 1982;30:961-81.
26. Loomes G, Sugden R. Regret theory: an alternative theory of rational choice. Economic J. 1982;92:805-24.
27. Zeelenberg M, Pieters R. A theory of regret regulation 1.1. J Consumer Psychol. 2007;17:29-35.

28. Velanovich V, Wollner I. Quality of life and performance status in patients with pancreatic and periampullary tumors. Int J Clin Oncol. 2011 Aug;16(4):401-7. PubMed PMID: 21347628. Epub 2011/02/25. eng.
29. Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. Stat Med. 2004 Mar 15;23(5):723-48. PubMed PMID: 14981672. Epub 2004/02/26. eng.
30. StatCorp. Stata Statistical Software: Release 12. College Station, TX: StataCorp LP; 2011.
31. Cohen SJ, Dobelbower R, Lipsitz S, Catalano PJ, Sischy B, Smith TJ, et al. A randomized phase III study of radiotherapy alone or with 5-fluorouracil and mitomycin-C in patients with locally advanced adenocarcinoma of the pancreas: Eastern Cooperative Oncology Group study E8282. International Journal of Radiation Oncology* Biology* Physics. 2005;62(5):1345-50.
32. Djulbegovic B, Hozo I, Lyman GH. Linking evidence-based medicine therapeutic summary measures to clinical decision analysis. MedGenMed. 2000;2(1):E6.
33. Hunink MM, Glasziou PP, Siegel JE, Weeks JC, Pliskin JS, Elstein AS, et al. Decision making in health and medicine: integrating evidence and values: Cambridge University Press; 2001.
34. Lichtenstein S, Slovic P. The construction of preference: Cambridge University Press; 2006.
35. Stiggelbout AM, De Haes J. Patient preference for cancer therapy: an overview of measurement approaches. Journal of Clinical Oncology. 2001;19(1):220-30.
36. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Medical Informatics and Decision Making. 2008;8(1):53.
37. Kahneman D. Maps of bounded rationality: Psychology for behavioral economics. American economic review. 2003:1449-75.
38. Birkmeyer JD, Finlayson SRG, Tosteson ANA, Sharp SM, Warshaw AL, Fisher ES. Effect of hospital volume on in-hospital mortality with pancreaticoduodenectomy. Surgery. 1999;125(3):250-6.
39. Gouma DJ, Van Geenen RCI, van Gulik TM, de Haan RJ, de Wit LT, Busch ORC, et al. Rates of complications and death after pancreaticoduodenectomy: risk factors and the impact of hospital volume. Annals of surgery. 2000;232(6):786.
40. Rudloff U, Maker AV, Brennan MF, Allen PJ. Randomized clinical trials in pancreatic adenocarcinoma. Surgical oncology clinics of North America. 2010;19(1):115.
41. Lal A, Christians K, Evans DB. Management of borderline resectable pancreatic cancer. Surgical oncology clinics of North America. 2010;19(2):359.
42. Katz MHG, Pisters PWT, Evans DB, Sun CC, Lee JE, Fleming JB, et al. Borderline resectable pancreatic cancer: the importance of this emerging stage of disease. Journal of the American College of Surgeons. 2008;206(5):833-46.
43. Coricelli G. The potential role of regret in the physician–patient relationship: Insights from neuroeconomics. 2008.
44. Djulbegovic B, Hozo I, Beckstead J, Tsalatsanis A, Pauker SG. Dual processing model of medical decision-making. BMC Medical Informatics and Decision Making. 2012;12(1):94.
45. Freemantle N. Are decisions taken by health care professionals rational? A non systematic review of experimental and quasi experimental literature. Health Policy. 1996;38(2):71-81.
46. Courvoisier DS, Agoritsas T, Perneger TV, Schmidt RE, Cullati S. Regrets associated with providing healthcare: Qualitative study of experiences of hospital-based physicians and nurses. PLoS One. 2011;6(8):e23138.

Table 1. Patient Demographics and SF-36 Scores. Values are the mean ± SEM unless otherwise indicated

Male : Female, n (%)	70 : 86 (45% : 55%)
Age (yr.),	65.9 ± 10
Stage: n (%)	
I	61 (39%)
II	68 (44%)
III	25 (16%)
0	2 (1%)
SF-36 Scores:[1]	
Physical functioning	55.2 ± 31
Role-physical	35.5 ± 44
Role-emotional	57.4 ±46
Bodily pain	55.5 ± 30
Pretreatment vitality	41.8 ± 24
Mental health	70.3 ± 21
Social functioning	60.8 ± 31
General health	60.7 ± 22
Patients undergoing resection, n (%)	85 (54%)
Patients receiving chemotherapy, n (%)	103 (66%)
Survival (mo.)	15.7 ± 25

[1]SF-36 Health Survey, rated from 0 to 100 on a Likert scale, with higher scores indicating better health or physical function (ref).

Table 2. Hazard ratio estimates of the prediction model

	Hazard Ratio	**P>|z|**	**[95% conf. interval]**
Stage	1.994865	0.001	1.326723- 2.999486
Pretreatment vitality	.9849276	0.030	.971512- .9985284
Role-physical	**.**9884022	0.005	.9803665- .9965038

Figure 1. Dual Visual Analog Scales. The DVAS are used for the elicitation of the decision maker's threshold probability. The questions depicted are case-specific.

Figure 2. Decision model for performing surgery on patients suffering from pancreatic adenocarcinoma.

s denotes the probability of survival, $S \pm$ denotes surgery or no surgery, $D \pm$ denotes death or no death, U_i are the utilities associated with each outcome and Rg is the regret associated with each action. For example, $Rg(S-, D+)$ is the regret associated with not performing a surgery for a patient who died within 7 months.

Figure 3.Overall survival of patients with pancreatic adenocarcinoma expressed as Kaplan-Meier survival and 95% confidence interval bands. Vertical bars (|) denote censored observations.

Figure 4. Regret DCA for the survival model constructed using Cox regression on three variables.

Dashed and dotted line denotes the decision to perform surgery; solid line denotes the decision not to perform surgery on any patient; dashed line denotes the use of the survival model to perform surgery. The optimal strategy is the action that results in the least amount of regret in case it is proven wrong. For threshold probabilities of 0-5%, the optimal strategy is to perform surgery on all patients, while for threshold probabilities greater than 5% the optimal strategy is to consult the survival model. For threshold probabilities between 80-87%, the regret curve associated with the prediction model is subject to noise associated to the error of the prediction model therefore, we assume that the prediction model remains the least regretful strategy.

Figure 5. Schematic Representation of Decision Model.

Natural History of Patients With Lung Cancer Without Treatment: A Systematic Review

ABSTRACT

Purpose: To conduct a systematic review and meta-analysis of the natural history of patients with confirmed diagnosis of lung cancer without active treatment.

Methods: Relevant studies were identified by search of MEDLINE (PubMed) and CENTRAL electronic databases and abstract proceedings up to June 2011. All prospective or retrospective studies assessing prognosis of lung cancer patients without treatment were eligible for inclusion. Data on mortality was extracted from all included studies and pooled proportion of mortality was calculated as a back-transform of the weighted mean of the transformed proportions, using the random-effects model.

Results: Seven cohort studies (4,418 patients) and 15 randomized controlled trials (1,031 patients) were included in the meta-analysis. All studies assessed mortality without treatment in patients with non-small cell lung cancer (NSCLC). The pooled proportion of mortality without treatment in cohort studies was 0.97 (95% CI: 0.96 to 0.99) and 0.96 in randomized controlled trials (95% CI: 0.94 to 0.98) over median study periods of 8 and 3 years, respectively. The pooled proportion of mortality was 0.97 (95% CI 0.96 to 0.98) when data from cohort and randomized controlled trials were combined. Test of interaction showed a statistically non-significant difference between subgroups of cohort and randomized controlled trials. Overall the studies were of moderate methodological quality.

Conclusion: Systematic evaluation of evidence on prognosis of NSCLC without treatment shows that mortality is very high. Although limited by study design, these findings provide the basis for future trials to determine optimal expected improvement in mortality with innovative treatments.

INTRODUCTION

Cancer is a major public health concern globally. It is the most frequent cause of death in economically developed countries.[1] Among all cancers, lung cancer is the leading cause of cancer deaths worldwide. [2] In the United States, approximately 221,130 new cases of lung cancer (14% of all cancer diagnoses) are expected in 2011 out of which 156,940 deaths (27% of cancer deaths) are estimated due to lung cancer.[3] Given the incurative nature of lung cancer, it is considered a terminal illness with a 5-year survival rate of approximately 16% .[3]

Patients diagnosed with terminal illness such as lung cancer confront several decisions related to management of the disease. Opting for treatment (e.g. chemotherapy, radiotherapy, or surgery) instead of palliation or vice versa is one such critical decision. Depending on the stage of the disease, potential benefits of anticancer therapy intended to palliate specific tumor-related symptoms may be at the expense of treatment-related harms and the inconvenience associated with undergoing treatment. Other times, palliative care (e.g. pain medications or low dose radiotherapy)[4] rather than anticancer therapy may be preferable. Informed decision related to management of a terminal disease thus requires accurate prognosis of the disease with or without treatment.

Briefly, prognosis refers to the likelihood of an individual developing a particular health outcome over a given period of time, based on the individual's clinical and non-clinical profile.[5] Accurate assessment of prognosis is key to informed decision making. For example, if a patient is diagnosed with a terminal illness such as lung cancer, a prognostic question of critical concern to the patient, family, and the physician is how long the patient is expected to live. Other important outcomes may include disease progression, health-related quality of life, and treatment-related harms. Reliable prognostication of life expectancy can prevent subjecting patients to costly and unnecessary treatment for an unduly long period before transitioning to hospice care.[6] This in turn can help patients

and their families prepare for the impending events and plan for the patient's remaining lifespan.[7] Accurate prognostic information can also help physicians decide on choice of curative versus palliative treatments. For instance, if evidence shows no effect of curative treatment on disease progression, significant treatment-related harms can be avoided in favor of palliative treatments.[7] Accurate disease prognosis thus underpins all management decisions related to the disease including choice of treatment, planning of supportive care, as well as allocation of resources.

Despite the significance of disease prognosis in clinical decision-making, systematic assessment of prognosis in patients with lung cancer without treatment has not been performed. We are aware of only one narrative review on the subject.[8] Accordingly, this systematic review was undertaken to assess the natural history of patients with confirmed diagnosis of lung cancer without active treatment. Specifically, our aim was to estimate overall survival (natural history) in lung cancer when no anticancer therapy is provided.

METHODS

This systematic review was conducted as per the methods elaborated in a protocol that was developed *a priori.* An ideal study design to assess natural history of a terminal disease such as lung cancer is a cohort study. Specifically, an inception cohort whereby a well-defined group of patients at the same disease stage is assembled at first diagnosis and followed for a defined period of time.[9-11] However, given the availability of treatments for lung cancer in recent years, it would be unethical and logistically challenging to conduct such a study. An alternative approach is to assess prognosis from retrospective lung cancer registries, case series or from the control arm of individual RCTs that compare active treatment with either no treatment or placebo or best supportive care.[5,12] Thus, in this review, any retrospective or prospective cohort study assessing prognosis in lung cancer without treatment and any RCT assessing the role of treatment versus no treatment, were eligible for inclusion. A study was eligible for inclusion irrespective of language or publication type.

Search Strategy

We conducted a systematic search of PubMed and Cochrane library electronic databases, proceedings of major scientific meetings, and bibliographies of eligible studies to identify all relevant studies. To retrieve lung cancer prognosis studies in PubMed, we employed search strategies suggested by Wilczynski[13] that optimizes search sensitivity and specificity. Search details used included: ("lung neoplasms"[MeSH Terms] AND "prognosis"[All Fields] AND "cohort"[All Fields] AND ("mortality"[Subheading] OR "natural course"[All Fields] OR "mortality"[All Fields] OR "survival"[All Fields] OR "survival"[MeSH Terms]).

To retrieve RCTs in PubMed, we employed strategies suggested by Haynes[14] with the following search details: ("lung neoplasms"[MeSH Terms] AND ("randomized controlled trial"[Publication Type]) AND ("palliative care"[All Fields] OR "hospice care"[All Fields] OR "supportive care"[All Fields] OR "best supportive care"[All Fields] OR "placebo"[All Fields] OR "symptomatic treatment"[All Fields] OR "no chemotherapy"[All Fields] OR "no treatment"[All Fields]).

In the Cochrane library, we utilized a free text search using the term “Lung cancer” to identify RCTs focusing on lung cancer. We manually searched abstracts of the American Society of Clinical Oncology (ASCO) and American Society of Hematology (ASH) meetings and utilized the snowballing procedure to identify other relevant studies. Studies published until June 2011 were included. No restrictions were made regarding the language of the publication.

Inclusion and Exclusion Criteria

A prospective or retrospective cohort study assessing overall survival as an outcome in lung cancer patients without treatment was eligible for inclusion. A RCT was included if it enrolled patients with confirmed diagnosis of lung cancer, compared treatment versus no treatment (e.g.

supportive care, best supportive care, palliative care, placebo etc.), and assessed overall survival as an outcome.

A study in which patients had anticancer treatment prior to enrollment and subgroup analyses were excluded. Additionally, RCTs comparing two active treatments were excluded. Two reviewers read the titles and abstracts of identified citations to identify potentially eligible studies. Full text of potentially relevant reports were retrieved and examined for eligibility. Disagreements about study inclusion or exclusion were resolved via discussion until a consensus was reached.

Data Extraction

Data extraction was performed using a standardized data extraction form. Two reviewers independently extracted the following information from each included study: number of patients enrolled, number of deaths, median survival, funding source (industry versus public etc.), type of centers involved (single versus multicenter etc.), patient demographics, patients baseline clinical characteristics, and type of control arm (for RCTs only). For cohort studies, we extracted data on the number of deaths and total number of patients diagnosed with lung cancer. For RCTs, we extracted data on the number of deaths (all-cause mortality) and number of participants randomized to the control arm.

Assessment of Methodological Quality

To evaluate the methodological quality of included studies, a modified checklist of predefined criteria was developed on four methodological domains pertinent to minimization of bias. This modified checklist uses applicable elements from existing tools (Quality in Prognosis Studies tool,[15] Evidence-Based Medicine Group criteria for prognostic studies,[16] Newcastle-Ottawa Quality Assessment Scale,[31] and Cochrane Collaboration risk of bias criteria[17]) and related studies (Hudak et al[18] and Altman[19]). The four domains included *participation bias* (extent to which study sample represents the population of interest on key characteristics), *attrition bias* (extent to which loss to

followup of the sample was not associated with key characteristics), *outcome measurement* (extent to which outcome of interest is adequately measured in study participants), *data analysis* and *reporting* (extent to which statistical analysis and data reporting are appropriate for the study design). The modified checklist contains 11 items for cohort studies and 14 items for RCTs. For each item, a study either fulfilled a certain criterion (scored "Yes") or failed to fulfill the criterion (scored "No"). To assess methodological quality of studies included, we focused on proportion of studies that fulfilled each quality criterion (Table 2).

Statistical Analysis

Data synthesis was conducted according to the study design separately as well as combined in the final stage (i.e., retrospective cohort and RCT).

For the purpose of meta-analysis, we used methods by Stuarts et al[20] to transform the proportions into a quantity according to the Freeman-Tukey variant of the arcsine square root transformed proportion. The pooled proportion was calculated as a back-transform of the weighted mean of the transformed proportions, using the random-effects model.

Heterogeneity of treatment effects between trials was assessed using the I^2 statistic[17] with the following thresholds for I^2 statistic values: low (25% to 49%), moderate (50% to 74%), and high (≥ 75%).[21] We explored the potential causes of heterogeneity by assessing the differences between subgroups using the test of interaction. We assessed robustness of the results by conducting sensitivity analysis with respect to methodological quality criteria of reporting, study location, and funding source. RevMan Version 5.1[22] was used to perform the analyses.

RESULTS

Literature Search

A flow diagram of the literature search is shown in Figure 1. Initial search identified 1,562 potentially relevant citations excluding 71 duplicates. After initial screening of titles and abstracts,

1,489 records were not relevant for reasons depicted in Figure 1 and were excluded. Further assessment of full texts of remaining 73 studies led to exclusion of 51 studies. Altogether, 22 studies met the pre-defined inclusion criteria: 7 were retrospective cohort studies[23-29] and 15 were RCTs.[30-44]

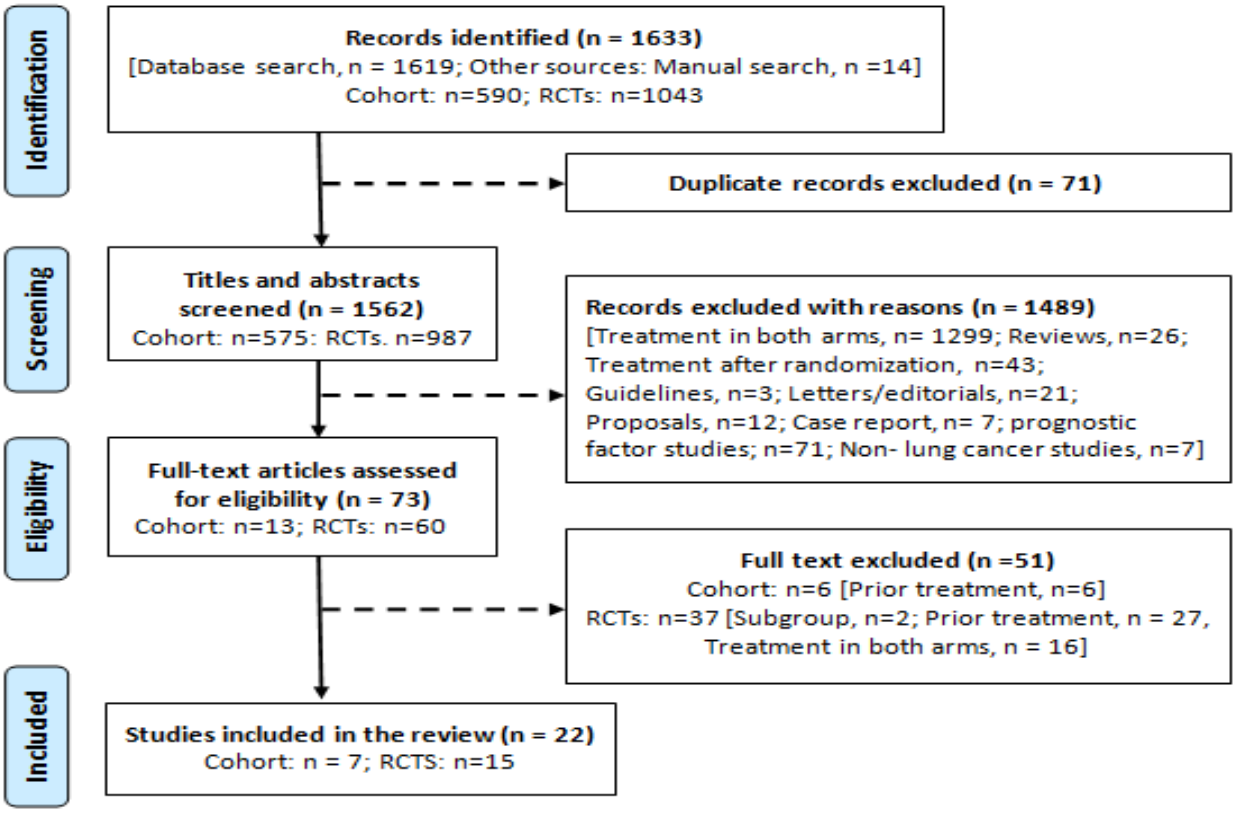


Figure 1 A flow diagram depicting the literature search process

Study Characteristics

We did not find any inception cohort study or a prospective cohort study assessing prognosis of patients with lung cancer without treatment. The seven retrospective cohort studies included 4,418 patients and the 15 RCTs enrolled 1,031 patients. Altogether, the 22 studies included 5,449 patients. All studies assessed prognosis in patients with NSCLC and were published between 1973 and 2009 (Table 1).

Cohort Studies: The median sample size in the cohort studies was 131 patients (range: 39 to 2,344 patients) with a median study period of 8 years (range: 5 to 13 years). Fifty-seven percent (4/7) and 29% (2/7) of the studies reported number of patients with stage I and stage II NSCLC,

respectively. Forty-three percent (3/7) of the studies reported patients' cancer histology. Seventy-one percent (6/7) of the studies reported patient's gender. Forty-three percent (3/7) of the studies reported median age. Forty-three percent (3/7) of the studies were conducted at single institutions, 43% (3/7) were at multicenter national studies, and 14% (1/7) of the studies had unspecified study location. Twenty-nine percent (2/7) of the studies were publicly funded, 14% (1/7) were funded by both public and industry, and 57% (4/7) had not specified funding sources.

RCTs: The median number of patients enrolled in the RCTs was 61 patients (range: 17 to 176 patients) with a median study period of 3 years (range: 1 to 7 years). Median follow-up was reported in 33% (5/15 of RCTs) and ranged between 2.7 and 43 months. Seventy-three percent (11/15) of the studies reported number of patients with stage III/IV NSCLC. Seventy-three percent (13/15) of the studies reported patients' cancer histology. Eighty-seven percent (13/15) of the RCTs reported patient's gender and median age. Twenty percent (3/15) of the RCTs were conducted at single institutions, 27% (4/15) were at multicenter national studies, 20% (3/15) were at multicenter international, and 33% (5/15) had unspecified study location. Seven percent (1/15) of the RCTs were funded by public, 33% (5/15) were funded by industry, 7% (1/15) were funded both public and industry, and 53% (8/15) had unspecified funding sources.

Types of control in RCTs: Three studies described *best supportive care* as comprising "symptomatic or palliative treatment excluding chemotherapy,"[45] "palliative radiotherapy, antibiotics, and corticosteroids,"[31] "palliative radiotherapy, opioid analgesics, and psychosocial support,"[38] or "radiation therapy, pain medication, nutritional and psychological support, thoracocentesis and/or tube thorascopy."[44] Three studies described *supportive care* as comprising "analgesics, an antitussive, relief of increased intracranial pressure, palliative radiotherapy, treatment of infections and pleural effusions,"[31] "symptomatic irradiation to involved fields,"[32] or "palliative radiation, analgesics, and psychosocial/nutritional support."[36] *Palliative care* consisted of "radiotherapy, antibiotics, coughs

suppressants, and analgesics"[34] *Symptomatic treatment* included "glucocorticosteroids and anabolic steroids." [39] No descriptions were provided for *placebo* and "*no treatment*."

Table 1 Characteristics of studies included in the review

Study	N	Study period (years)	Disease Stage		Histology			Male	Median Age (years)
			I	II	squamous	adeno	large-cell		
(a) Cohort studies									
Raz 2007	1432	13	1432	NR	460	419	89	747	74
Wisnivesky 2007†	2344	8	NR	NR	NR	NR	NR	1292	NR
Chadha 2005	39	11	23	13	18	88	5	4	77
Henschke 2003	131	7	131	NR	NR	NR	NR	NR	NR
McGarry 2002†	49	5	NR	NR	NR	NR	NR	49	NR
Vrdoljak 1994	130	7	55	56	61	35	34	120	60
Hyde 1973	293	8	NR	NR	NR	NR	NR	NR	NR
Total/[Range]	4418	[5-13]	1641	68	539	542	128	2211	
(b) RCTs			**III**	**IV**					
Goss 2009[m]	101	2 [0.23]	17	84	25	46	11	61	76
Anderson 2000	150	2	92	58	NR	NR	NR	91	64
ELVIS 1999 [m]	78	1 [1.08]	22	56	33	29	3	69	74*
Cullen 1999 [m]	176	8 [2.17]	88	88	103	42	6	122	64
Thongprasert 1999	98	4	49	49	31	49	12	NR	60
Helsing 1998 [m]	26	5 [3.33]	3	23	5	17	4	18	65
Cartei 1993	50	7	NR	50	25	17	8	36	57
Leung 1992 [m]	66	4 [3.58]	58	NR	31	18	7	48	62
Cellerino 1991	61	3	61	NR	38	18	5	59	62
Quoix 1991	22	3	NR	22	NR	NR	NR	NR	NR
Kaasa 1991	43	3	NR	43	16	16	11	31	62*
Ganz 1989	26	2	NR	26	9	17	NR	23	NR
Rapp 1988	50	3	50	NR	12	24	12	38	58
Cormier 1982	17	2	17	NR	8	2	6	16	60
Laing 1975	67	2	15	20	23	5	9	59	64
Total/[Range]	1031	[1-8]	472	519	359	300	94	671	[57-76]

Note: **N** = Sample size or number of participants enrolled; **NR**= data not reported; † = Sample includes stage I and II cancer; **adeno** = adenocarcinoma; **squamous** = squamous cell carcinoma; **large-cell** = large-cell carcinoma; *=we recorded mean age where median age was not reported or not extractable, **[m]** = median follow-up in parenthesis

Methodological Quality

Cohort: All seven cohort studies fulfilled 64% (7/11) of the quality criteria (Table 2). That is, adequate description of population of interest for key characteristics, adequate description of study setting/geographic location, adequate participation in the study by all eligible patients, reporting of patients with missing data, a priori and objective definition of outcomes, and presentation of frequencies of most important data (e.g., outcome) were reported in all studies. However, baseline sample was adequately described for key characteristics in 57% (4/7) of the studies, inclusion and

exclusion criteria were adequately described in 71% (5/7) of the studies, follow-up was sufficiently long for outcome to occur in 86% (6/7) of the studies, and alpha error and/or beta error were specified *a priori* in 29% (2/7) of the studies.

RCTs: All 15 RCTs fulfilled 36% (5/14) of the quality criteria (Table 2). That is, adequate description of population of interest for key characteristics, adequate description of withdrawal (incomplete outcome data), a priori and objective definition of outcomes, and frequencies of most important data were reported in all RCTs. However, study setting and geographic location were adequately described in 47% (7/15) of the RCTs, baseline sample was adequately described for key characteristics in 93% (14/15) of the RCTs, inclusion and exclusion criteria were adequately described in 93% (14/15) of the RCTs, patients were balanced in all aspects except the intervention in 93% (14/15) of the RCTs, follow-up was sufficiently long for outcome to occur in 53% (8/15) of the RCTs, proportion of sample completing the study was adequate in 60% (9/15) of the RCTs, characteristics of dropouts versus completers was provided in 13% (2/15) of the RCTs, alpha error and/or beta error was specified a priori in 47% (7/15) of the RCTs, and data analysis was based on intention to treat analysis principle in 53% (9/15) of the RCTs.

Table 2 Methodological Quality of Lung Cancer prognosis Studies

Study Design/Domain/Criterion		Criteria fulfilled	
		n/N	%
Cohort studies (11 items)			
	Participation bias		
A	Population of interest is adequately described for key characteristics[15]	7/7	100
B	Study setting and geographic location is adequately described[15]	7/7	100
C	Baseline sample is adequately described for key characteristics[15]	4/7	57
D	Inclusion and exclusion criteria are adequately described[15]	5/7	71
E	There is adequate participation in the study by all eligible patients[15]	7/7	100
	Attrition bias		
F	Follow-up is sufficiently long for outcome to occur (≥ 6 months)[16,18,19,46]	6/7	86
G	Patients with missing data were reported[15,17]	7/7	100
	Outcome measurement		
H	Definition of outcome is provided *a priori* [15]	7/7	100
I	Objective definition of outcome is provided[15,16,18,19]	7/7	100
	Data analysis and reporting		
J	Alpha error and/or beta error is specified *a priori*	2/7	29

K	Frequencies of most important data (e.g., outcomes) are presented[18,19,47]	7/7	100
Randomized Controlled Trials (15 items)			
	Participation bias		
L	Population of interest is adequately described for key characteristics[15]	15/15	100
M	Study setting and geographic location is adequately described[15]	7/15	47
N	Baseline sample is adequately described for key characteristics[15]	14/15	93
O	Inclusion and exclusion criteria are adequately described[15]	14/15	93
P	Patients were balanced in all aspects except the intervention	15/15	93
	Attrition bias		
Q	Follow-up is sufficiently long for outcome to occur (≥ 6 months)[16,18,19,46,48]	8/15	53
R	Proportion of sample completing the study is adequate (≥80%)[15,16,18,47,49,50]	9/15	60
S	Description of withdrawal (incomplete outcome data) is provided[15,17]	15/15	100
T	Characteristics of dropouts versus completers is provided[15]	2/15	13
	Outcome measurement		
U	Definition of outcome is provided *a priori*[15]	15/15	100
V	Objective definition of outcome is provided[15,16,18,19]	15/15	100
	Data analysis and reporting		
W	Alpha error and/or beta error is specified *a priori*	7/15	47
X	Data analysis was based on intention to treat analysis principle[17]	9/15	53
Y	Frequencies of most important data (e.g., outcomes) are presented[18,19,47]	15/15	100

Mortality

Cohort: Data on mortality was extractable from all seven cohort studies enrolling 4,418 patients. As shown in Figure 2, the pooled proportion of mortality for patients without anticancer treatment was 0.97 (95%CI: 0.96 to 0.99). There was a statistically significant heterogeneity among pooled cohort studies (I^2 =93%, P < 0.00001).

RCTs: Data on mortality was extractable from the control arm of all 15 RCTs (1,031 patients). Figure 2 shows that the pooled proportion of mortality for patients in the control arm (without active treatment) was 0.96 (95% CI: 0.94 to 0.98). There was a statistically significant heterogeneity among pooled control arm of RCTs (I^2 =80%, P < 0.00001).

Combined (Cohort and RCTs): Pooled proportion of mortality across the 22 studies was 0.97 (95%CI: 0.96 to 0.98). Because these two designs are inherently different from each other, we conducted separate analyses. However, as shown in Figure 2, test for subgroup differences showed no statistically significant heterogeneity between the two study designs (P = 0.28).

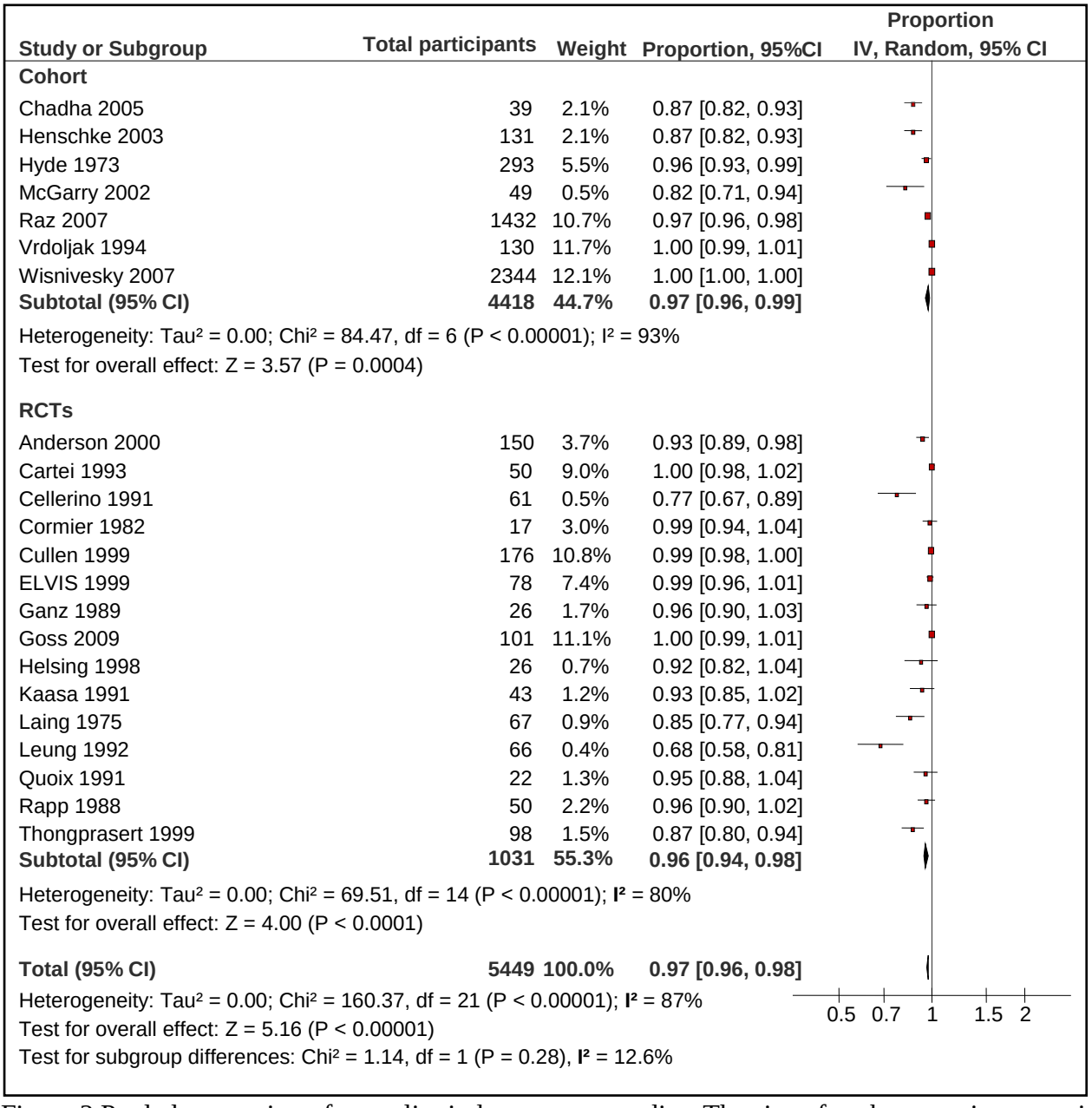


Figure 2 Pooled proportion of mortality in lung cancer studies. The size of each square is proportional to the weight of the study (inverse variance)

Sensitivity Analysis

To assess the robustness of overall results according to the study design (cohort vs. RCT) as well as explore the reasons for observed heterogeneity in the pooled proportion of mortality, we conducted additional sensitivity analyses. For both cohort studies and RCTs, we conducted sensitivity analyses according to methodological quality criteria, funding source, and study location. For RCTs only, we conducted additional sensitivity analyses according to type of control. The results of sensitivity

analyses are summarized in Figure 3. Overall, the results remained unchanged in the sensitivity analyses. There were no statistically significant differences in the proportion of mortality.

Cohort: In cohort studies, there was no statistically significant difference in the proportion of mortality according to any methodological criteria of reporting. With respect to study location, the pooled proportion of mortality in cohort studies conducted at multicenter national locations was 0.95 (95%CI: 0.89 to 1.01) and at single institution was 0.98 (95%CI: 0.95 to 1.01) whereas the pooled proportion of mortality in cohort studies conducted at unspecified locations was 0.87 (95%CI: 0.82 to 0.93). Test for overall interaction among these subgroups was statistically significant (P = 0.007). Regarding funding source, the pooled proportion of mortality in public-funded, unspecified funding sources, and public/industry-funded cohort studies were 1.00 (95%CI: 1.00 to 1.00), 1.00 (95%CI: 0.99 to 1.00), and 0.97 (95%CI: 0.96 to 0.98), respectively. The test for overall interaction among these subgroups was statistically significant (P < 0.0001).

RCTs: There was no statistically significant difference in the proportion of mortality according to methodological criteria of reporting, study location, and funding source. With respect to type of control, the pooled proportion of mortality in RCTs involving best supportive care, no treatment, placebo, supportive care, and symptomatic treatment as control were 0.90 (95%CI: 0.83 to 0.97) and in RCTs involving supportive care as control was 0.96 (95%CI: 0.92 to 1.00), 0.86 (95%CI: 0.81 to 0.92), 1.00 (95%CI: 0.99 to 1.01), 0.96 (95%CI: 0.92 to 1.00), and 0.97 (95%CI: 0.92 to 1.03), respectively. Test for overall interaction among these subgroups was statistically significant (P < 0.00001).

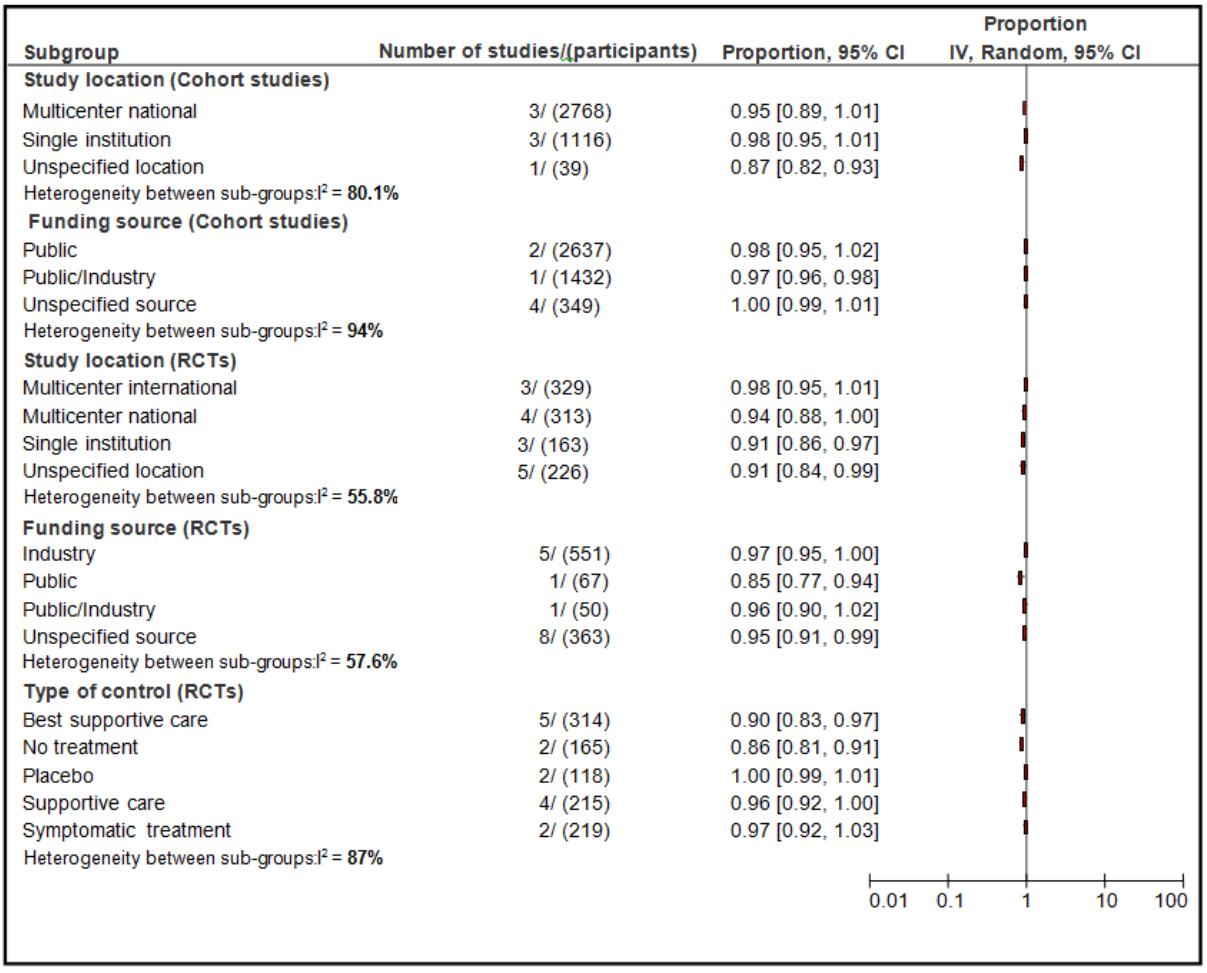


Figure 3 Pooled Proportions of Mortality and Heterogeneity Between Subgroups

DISCUSSION

This is the first study to provide most comprehensive data related to natural history of lung cancer. The results show that prognosis of patients with lung cancer not receiving treatment is very high. Regardles of the study design (i.e. cohort versus RCTs) the findings were similar and did not differ according to disease severity. For example, all cohort studies assessed mortality in patients with early stage NSCLC (stage I/II) and all RCTs enrolled patients with advance stage NSCLC (stage III/IV). However, the mortality rates from cohort and RCTs essentially remained unchanged (97% vs 96%). Overall, included studies were of moderate methodological quality.

The findings from our study is similar to the study by Detterbeck and Gibson[4] which showed a 98% 5-year mortality rate for stage I/II lung cancer (median survival = 10 months). Despite the obvious similarity in results our study is significantly different in the conduct and analysis. For example, the study by Detterbeck and Gibson[4 4] did not employ a systematic approach to data collection and analysis (i.e. not a systematic review) and therefore the findings are not reproducible. The similarity in findings might be an artifact of play of chance. Furthermore, quantitative synthesis of results across included studies was not performed in the study by Detterbeck and Gibson[4] which was undertaken in our study. Another unique feature of our study lies in the inclusion of RCTs in addition to retrospective studies. None of the previous studies on the topic have utilized the approach of pooling data from one arm of RCTs for accurate assessment of prognosis. Therefore, due to the reasons enumerated here the study presented here is the most comprehensive to date reporting the natural history of lung cancer.

Our study has some limitations. For example, we observed a statistically significant heterogeneity in pooled results which we could not explain through subgroup analyses. We suspect that the observed heterogeneity is clinical and not methodological. Specifically in the case of RCTs, the constitution of control arm varied across pooled studies. For example, five RCTs employed best supportive care as control, four had supportive care, two had placebo, two had no treatment and another two had symptomatic treatment as control. While, the definitions are very clear on placebo and no treatment, which was also explained by the sensitivity analyses (I^2 =0% for both subgroups), the composition of best supportive care, supportive care, and symptomatic treatment varied significantly across pooled studies. In these cases, the observed heterogeneity remained unexplained. The findings are also limited in terms of generalizability by the fact that all included studies enrolled patients with NSCLC due to which the results are not entirely applicable to all lung cancers. However,

it is important to note that a systematic review is limited by the availability of data and we did include all available data related to prognosis of lung cancer patients without treatment.

Comprehensive data on the natural history of lung cancer is required for informed decision making by patients, physicians and researchers. For patients, it serves as the basis for their expected outcome with and without treatment, which is critical in cases of diseases with high mortality. For physicians, accurate and reliable information facilitates shared decision making with patients related to choice of interventions or no intervention. Most importantly, the findings are needed by researchers to avoid *optimism bias*.[51] Briefly, optimism bias refers to unwarranted belief in the efficacy of new therapies. A study by Djulbegovic et al. [51] assessed the role of optimism bias in a cohort of trials conducted by the National Cancer Institute Cooperative Groups and concluded that the optimism bias is the primary reason for inconclusive findings in the context of RCTs. Accordingly, the results from our study will help researchers determine the most optimal rate of expected improvement in mortality with innovative/newer treatments.

Funding

Department of Army funding was provided to the third author to develop computer decision-support system for better prognostication in life expectancy and improvement in decision-making in terminally ill patients. Department of Army grant #W81 XWH 09-2-0175 (PI: B. Djulbegovic)

REFERENCES

1. Jemal A. Global Cancer Statistics (vol 61, pg 69, 2011). *Ca-Cancer J Clin.* Mar-Apr 2011;61(2):134-134.

2. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer.* Dec 15 2010;127(12):2893-2917.

3. Society AC. *Cancer Facts and Figures 2011*2011.

4. Detterbeck FC, Gibson CJ. Turning gray: the natural history of lung cancer over time. *J Thorac Oncol.* Jul 2008;3(7):781-792.

5. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? Vol 338.

6. Christakis NA, Lamont EB. Extent and determinants of error in doctors' prognoses in terminally ill patients: prospective cohort study. *Bmj.* Feb 19 2000;320(7233):469-472.

7. Mackillop WJ, Quirt CF. Measuring the accuracy of prognostic judgments in oncology. *J Clin Epidemiol.* Jan 1997;50(1):21-29.

8. Daugherty CK, Hlubocky FJ. What Are Terminally Ill Cancer Patients Told About Their Expected Deaths? A Study of Cancer Physicians' Self-Reports of Prognosis Disclosure. *J Clin Oncol.* Dec 20 2008;26(36):5988-5993.

9. Hemingway H, Riley RD, Altman DG. Ten steps towards improving prognosis research. Vol 339.

10. Hemingway H. Prognosis research: why is Dr. Lydgate still waiting? *J Clin Epidemiol.* Dec 2006;59(12):1229-1238.

11. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *Bmj.* 2009;338:b375.

12. Chalmers I, Glasziou P, Greenhalgh T, et al. Table: Steps in finding evidence ("Level") for different types of question.

13. Wilczynski NL, Haynes RB. Developing optimal search strategies for detecting clinically sound prognostic studies in MEDLINE: an analytic survey. *BMC Med.* Jun 9 2004;2:23.

14. Haynes RB, McKibbon KA, Wilczynski NL, Walter SD, Werre SR. Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytical survey. *Bmj.* May 21 2005;330(7501):1179.

15. Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. *Ann Intern Med.* Mar 21 2006;144(6):427-437.

16. Laupacis A, Wells G, Richardson WS, et al. Users Guides to the Medical Literature .5. How to Use an Article About Prognosis. *Jama-J Am Med Assoc.* Jul 20 1994;272(3):234-237.

17. Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2 [updated September 2009]. The Cochrane Collaboration, 2009. Available from www.cochrane-handbook.org.

18. Hudak PL, Cole DC, Frank JW. Perspectives on prognosis of soft tissue musculoskeletal disorders. *Int J Rehabil Res.* Mar 1998;21(1):29-40.

19. Altman DG. Systematic reviews of evaluations of prognostic variables. *Bmj.* Jul 28 2001;323(7306):224-228.

20. Stuart A. Kendall's advanced theory of statistics. Jan 5 1994;102(1):173-180.

21. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Vol 3272003:557-560.

22. *Review Manager (RevMan)* [computer program]. Version 5.0. Copenhagen: The Nordic Cochrane Centre; 2008.

23. Chadha AS, Ganti AK, Sohi JS, Sahmoun AE, Mehdi SA. Survival in untreated early stage non-small cell lung cancer. *Anticancer Res.* Sep-Oct 2005;25(5):3517-3520.

24. Henschke CI, Wisnivesky JP, Yankelevitz DF, Miettinen OS. Small stage I cancers of the lung: genuineness and curability. *Lung Cancer.* Mar 2003;39(3):327-330.

25. Hyde L, Wolf J, McCracken S, Yesner R. Natural course of inoperable lung cancer. *Chest.* Sep 1973;64(3):309-312.

26. McGarry RC, Song G, des Rosiers P, Timmerman R. Observation-only management of early stage, medically inoperable lung cancer: poor outcome. *Chest.* Apr 2002;121(4):1155-1158.

27. Raz DJ, Zell JA, Ou SH, Gandara DR, Anton-Culver H, Jablons DM. Natural history of stage I non-small cell lung cancer: implications for early detection. *Chest.* Jul 2007;132(1):193-199.

28. Vrdoljak E, Mise K, Sapunar D, Rozga A, Marusic M. Survival analysis of untreated patients with non-small-cell lung cancer. *Chest.* Dec 1994;106(6):1797-1800.

29. Wisnivesky JP, Halm EA. Sex differences in lung cancer survival: do tumors behave differently in elderly women? *J Clin Oncol.* May 1 2007;25(13):1705-1712.

30. Anderson H, Hopwood P, Stephens RJ, et al. Gemcitabine plus best supportive care (BSC) vs BSC in inoperable non-small cell lung cancer - a randomized trial with quality of life as the primary outcome. *Br J Cancer.* 2000;83(4):447-453.

31. Cartei G, Cartei F, Cantone A, et al. Cisplatin-cyclophosphamide-mitomycin combination chemotherapy with supportive care versus supportive care alone for treatment of metastatic non-small-cell lung cancer. *J Natl Cancer Inst.* May 19 1993;85(10):794-800.

32. Cellerino R, Tummarello D, Guidi F, et al. A randomized trial of alternating chemotherapy versus best supportive care in advanced non-small-cell lung cancer. Vol 91991:1453-1461.

33. Cormier Y, Bergeron D, La Forge J, et al. Benefits of polychemotherapy in advanced non-small-cell bronchogenic carcinoma. *Cancer.* Sep 1 1982;50(5):845-849.

34. Cullen MH, Billingham LJ, Woodroffe CM, et al. Mitomycin, Ifosfamide, and Cisplatin in Unresectable Non–Small-Cell Lung Cancer: Effects on Survival and Quality of Life. Vol 171999:3188-3194.

35. Elderly Lung Cancer Vinorelbine Italian Study Group T. Effects of Vinorelbine on Quality of Life and Survival of Elderly Patients With Advanced Non-Small-Cell Lung Cancer. Vol 911999:66-72.

36. Ganz PA, Figlin RA, Haskell CM, La Soto N, Siau J. Supportive care versus supportive care and combination chemotherapy in metastatic non-small cell lung cancer. Does chemotherapy make a difference? *Cancer.* Apr 1 1989;63(7):1271-1278.

37. Goss G, Ferry D, Wierzbicki R, et al. Randomized Phase II Study of Gefitinib Compared With Placebo in Chemotherapy-Naive Patients With Advanced Nonâ€“Small-Cell Lung Cancer and Poor Performance Status. Vol 272009:2253-2260.

38. Helsing M, Bergman B, Thaning L, Hero U. Quality of life and survival in patients with advanced non-small cell lung cancer receiving supportive care plus chemotherapy with carboplatin and etoposide or supportive care only. A multicentre randomised phase III trial. *European Journal of Cancer.* 1998;34(7):1036-1044.

39. Kaasa S, Lund E, Thorud E, Hatlevoll R, Host H. Symptomatic treatment versus combination chemotherapy for patients with extensive non-small cell lung cancer. *Cancer.* May 15 1991;67(10):2443-2447.

40. Laing AH, Berry RJ, Newman CR, Peto J. TREATMENT OF INOPERABLE CARCINOMA OF BRONCHUS. *The Lancet.* 1975;306(7946):1161-1164.

41. Leung WT, Shiu WC, Pang JC, et al. Combined chemotherapy and radiotherapy versus best supportive care in the treatment of inoperable non-small-cell lung cancer. *Oncology.* 1992;49(5):321-326.

42. Quoix E, Dietemann A, Charbonneau J, et al. [Is chemotherapy with cisplatin useful in non small cell bronchial cancer at staging IV? Results of a randomized study]. *Bull Cancer.* 1991;78(4):341-346.

43. Rapp E, Pater JL, Willan A, et al. Chemotherapy can prolong survival in patients with advanced non-small- cell lung cancer--report of a Canadian multicenter randomized trial. Vol 61988:633-641.

44. Thongprasert S, Sanguanmitra P, Juthapan W, Clinch J. Relationship between quality of life and clinical outcomes in advanced non-small cell lung cancer: best supportive care (BSC) versus BSC plus chemotherapy. *Lung Cancer.* 1999;24(1):17-24.

45. Anderson G, Payne H. Response rate and toxicity of etoposide (VP-16) in squamous carcinoma of the lung: report from the Lung Cancer Treatment Study Group. *Semin Oncol.* Mar 1985;12(1 Suppl 2):21-22.

46. Kuijpers T, van der Windt DA, van der Heijden GJ, Bouter LM. Systematic review of prognostic cohort studies on shoulder disorders. *Pain.* Jun 2004;109(3):420-431.

47. Nijrolder I, van der Horst H, van der Windt D. Prognosis of fatigue. A systematic review. *J Psychosom Res.* Apr 2008;64(4):335-349.

48. Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *Eur J Epidemiol.* Sep 2010;25(9):603-605.

49. Cole MG, Ciampi A, Belzile E, Zhong L. Persistent delirium in older hospital patients: a systematic review of frequency and prognosis. *Age Ageing.* Jan 2009;38(1):19-26.

50. Hermens ML, van Hout HP, Terluin B, et al. The prognosis of minor depression in the general population: a systematic review. *Gen Hosp Psychiatry.* Nov-Dec 2004;26(6):453-462.

51. Djulbegovic B, Kumar A, Magazin A, et al. Optimism bias leads to inconclusive results-an empirical study. *Journal of Clinical Epidemiology.* Jun 2011;64(6):583-593.

29(3) MS 2 Miladinovic; 37Rs, 2Ts, 2Fs
[running head: External Validation of a Web-Based Prognostic Tool]

External Validation of a Web-Based Prognostic Tool for Predicting Survival for Patients in Hospice Care

Branko Miladinovic, Rahul Mhaskar, Ambuj Kumar, Sehwan Kim, Ronald Schonwetter, and Benjamin Djulbegovic
B Miladinovic (corresponding author): Center for Evidence-Based Medicine and Health Outcomes Research, Morsani College of Medicine, University of South Florida, 3515 E. Fletcher Avenue, MDT1200, Mail Code MDC27, Tampa, Florida 33612, USA; bmiladin@health.usf.edu

R Mhaskar, A Kumar: Center for Evidence-Based Medicine and Health Outcomes Research, University of South Florida, Tampa, Florida, USA; **S Kim, R Schonwetter**: HPC Healthcare, Temple Terrace, Florida, USA; **B Djulbegovic**: Center for Evidence-Based Medicine and Health Outcomes Research, University of South Florida, and H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida, USA

Keywords: Prognostat, survival prognostication, Palliative Performance Scale, external validation, Royston-Parmar survival models

Abstract / Prognostat is an interactive Web-based prognostic tool for estimating hospice patient survival based on a patient's Palliative Performance Scale (PPS) score, age, gender, and cancer status. The tool was developed using data from 5,893 palliative care patients, which was collected at the Victoria Hospice in Victoria, British Columbia, Canada, beginning in 1994. This study externally validates Prognostat with a retrospective cohort of 590 hospice patients at LifePath Hospice and Palliative Care in Florida, USA. The criteria used to evaluate the prognostic performance were the Brier score, area under the receiver operating curve, discrimination slope, and Hosmer-Lemeshow goodness-of-fit test. Though the Kaplan-Meier curves show each PPS level to be distinct and significantly different, the findings reveal low agreement between observed survival in our cohort of

patients and survival predicted by the prognostic tool. Before developing a new prognostic model, researchers are encouraged to update survival estimates obtained using Prognostat with the information from their cohort of patients. If it is to be useful to patients and clinicians, Prognostat needs to explicitly report patient risk scores and estimates of baseline survival.

INTRODUCTION

Accurate prognostication of hospice patient survival gives patients and their family members a vital opportunity to attend to matters such as planning, prioritizing, and preparing for death (1). Predicting patient survival without using a prognostic model is often affected by optimism or avoidance, which can lead to poor prediction of life expectancy. Studies have shown that clinicians consistently overestimate survival times of terminally ill patients (2-4). One prospective cohort study suggested that doctors overestimated survival of terminally ill patients by a factor of 5 (5). Successful prognostication of patient survival depends on developing and testing prognostic models, which entails having accurate patient data for prognosis and selecting clinically relevant candidate predictors and measures of model performance, usually in the context of a multivariable regression survival model (6). This process produces patient performance scores that allow for classification of patients into different risk groups.

The usefulness and validity of a prognostic model are judged by how well the model performs for patients who come from different centres (7). A validated prognostic model is generally accepted to be one that works in a data set other than the one that has been used to develop it (7, 8). There is also a general concurrence that the validation process should follow guidelines and that unvalidated prognostic models should not be applied in clinical practice (9-11). As the value of any prediction model is its generalizability to other groups of patients, our goal was to externally validate Prognostat (12) — a Web-based interactive prognostic tool for estimating hospice patient survival — on a

retrospective cohort of 590 hospice patients in Florida, USA. Prognostat estimates survival times based on palliative patients' age group, gender, diagnosis, and score on the Palliative Performance Scale (PPS) (13).

In this paper, we discuss Prognostat and introduce the measures of model performance. Since predictive performance may decrease when Prognostat is tested with new patients as compared to the patients who were used to develop the model, we also discuss a strategy for updating Prognostat in future studies.

METHODS

Study Sample and Survival Estimation Using Prognostat

The patient data were obtained from LifePath Hospice and Palliative Care, licensed since 1983 to serve Hillsborough County, Florida. The data for 590 consecutive deceased patients was extracted starting in January 2009 and working backwards. This study was a retrospective review of deceased patients' medical records, and only data that pertained to outcomes was collected; personal information was not collected, and data were de-identified prior to analysis. A trained nurse assigned PPS scores at admission to our cohort of patients. The University of South Florida's institutional review board approved the study. Two research assistants extracted all data necessary to populate the model variables, and two faculty members (RM and BD) randomly checked 25 percent of the data for accuracy.

Prognostat was developed at the University of Victoria (in Victoria, British Columbia, Canada) using retrospective survival estimates of 5,893 palliative care patients collected at the Victoria Hospice starting in 1994. It calculates survival rate in days for the variables or covariates found to be statistically significant predictors of patient survival — namely, the patient's gender, age group (19 to 44, 45 to 64, 65 to 74, 75 to 84, or 85 and over), diagnosis (lung cancer, breast cancer, colorectal cancer, prostate cancer, other cancer, or noncancer illness), and PPS score.

Decisions regarding hospice admission depend on the care an individual requires and the specific hospice setting. While US Medicare guidelines state that only individuals with a life expectancy of six months or less may be admitted to hospice in the US, the criteria for hospice admission in Canada vary among geographical areas and among individual hospices — that is, some Canadian hospices admit patients with a life expectancy of one month or less, while others do not impose such restrictions. Palliative care providers or programs will often assist patients in determining the best timing for admission to hospice.

The PPS was developed and reported by Anderson and colleagues (13) to measure the functional status of patients receiving palliative care. The scale has 11 possible mutually exclusive levels, from 0 (the patient is dead) to 100 (the patient is ambulatory and healthy). Numerous studies have assessed its performance in a variety of settings and found it to be a statistically significant risk score for calculating survival estimates (14-22).

Prognostat survival estimates were derived using the Cox proportional hazards (CPH) model, which relies on both the baseline survival function and risk scores to estimate patient survival. Because reporting the baseline function under CPH is not possible and Prognostat does not explicitly report prognostic indices (or risk factors), it makes model calibration in other populations unfeasible.[i]

Assessment of Model Performance

Using measures of accuracy, discrimination, and calibration, we analyzed Prognostat's predictive performance based on the ability of the estimated risk score to predict survival. Accuracy refers to the difference between the probability of survival predicted with Prognostat and observed patient survival. The Brier score is a quadratic scoring rule that calculates the differences between actual outcomes and predicted probabilities (23). Given the predicted probability of survival p_i at time t for patient i, and Y_i binary (0-1, dead-alive) variable, the Brier score is defined as $\sum_i(Y_i\ (1\ -\ p_i)^2\ +\ (1\ -\ Y_i)\ p_i{}^2)$. A Brier score of 0 indicates a perfect model, while 0.25 indicates a non-informative model (the value

achieved when issuing a predicted probability of 50% to each patient). The Brier score may be scaled by its maximum $\text{Brier}_{max} = (1 - \text{mean}(p_i))\ \text{mean}(p_i)$ to obtain $\text{Brier}_{scaled} = \left(1 - \frac{\text{Brier}}{\text{Brier}_{max}}\right) 100\%$ which has interpretation similar to the Pearson correlation coefficient (24).

Calibration refers to how closely the predicted survival calculated at a pre-specified time using Prognostat agrees with the observed survival. Since calibration is essentially a test of fit, we applied the Hosmer-Lemeshow (HL) test (25) on the dead versus alive binary outcome. The HL Chi-square statistic involves grouping of the observations (most commonly in deciles) based on the predicted probabilities and then testing the hypothesis that the difference between observed and predicted events is simultaneously zero for all the groups. This test is equivalent to testing the hypothesis that the observed number of events in each of the groups is equal to the expected number of events based on the fitted model. The higher the HL p-value, the better calibrated the model is. The HL calibration can be visually expressed by plotting deciles of predicted versus observed proportions of survival at each time point.

Discrimination is the ability of the model to differentiate between the patients who died versus those who survived at a pre-specified time. A rank order statistic commonly used to summarize discrimination with and without the outcome has been the area under the receiver operating curve (AUC) (26), which is a plot of the sensitivity (true positive rate) against 1-specificity (false positive rate) for consecutive cutoffs of the probability of an outcome. The maximum value of the area under the receiver operating curve (AUC), AUC=1, indicates a perfect prediction model, while a value of AUC=0.5 indicates that 50 percent of patients have been correctly classified (as good as by chance). As a rank order statistic, AUC is insensitive to errors such as difference in average survival. For this reason, a model can have relatively moderate AUC scores and at the same time be inaccurate and have high Brier scores (or low-scaled Brier scores).

The discrimination slope is a measure of how well subjects with and without the outcome are separated. It is defined as the absolute difference in mean predictions of survival (mean [p_i]) between those who died and those who survived at time t (8). Because it is an overall measure of differences in mean survival probabilities, in addition to the discrimination slope we have used box plots to assess the extent to which survival differentiation at each time point is achieved for all survival estimates. All statistical calculations were performed using Stata version 11.2.

RESULTS

Patient characteristics of the retrospective cohort are summarized in Table 1. The extracted data were found to be in substantial agreement (kappa=0.85). In addition to presenting data for our cohort of 590 patients, in each column, as a second cell entry, we present data from the Victoria Hospice cohort that was used to develop Prognostat. The table shows significant discrepancies in the distribution of percentages for age and cancer status. There is also a significant discrepancy in the distribution of percentages and median survival times for PPS.

For our cohort, the Kaplan-Meier curves stratified by initial PPS level are shown in Figure 1. The curves show good separation, indicating that the different risk groups are well defined. We dropped 15 patients with PPS scores of 60 percent due to the crossing of the Kaplan-Meier estimate of PPS 50 percent. The log-rank test for equality of survival curves was highly significant at p=0.001 for PPS and cancer status, but not for age (p=0.303) and gender (p=0.944). Likewise, when adjacent categories of PPS were compared (PPS 10 percent versus 20 percent, 20 percent versus 30 percent, and so on), pairwise log-rank tests were all significant at p=0.05 level, except for PPS 40 percent versus PPS 50 percent (p=0.394), due to initial crossing of two survival curves and the longer tail of the PPS 40 percent group. Patients who were 44 years old or younger did not have significantly lower hazard than those in the other age groups (p=0.862, 0.340, 0.466, 0.50, respectively), nor did male patients compared with female ones (p=0.806).

The measures of accuracy, discrimination, and calibration for days 1, 2, 4, 7, 14, and 30 are given in Table 2 and show poor performance of Prognostat overall. The discrimination slopes are relatively low and the Hosmer-Lemeshow (HL) goodness-of-fit test *p*-values are significant for all six days of measurement, indicating poor calibration. In the HL calibration plot of predicted versus observed proportion of those who survived (Figure 2B), circles are mostly unaligned with the 45-degree line. They show that in our cohort of patients, Prognostat consistently underestimates survival for days 1, 2, 4, 7, and 14, and overestimates it for day 30. The larger circles indicate that these points are based on more data. The absence of circles in any given decile indicates that there were no predictions in that interval. The overlapping box plots (Figure 2A) confirm poor discrimination.

DISCUSSION

This paper describes an external validation of the Web-based interactive prognostic tool Prognostat. We found that the tool performed poorly for our cohort of palliative patients. Since patient populations differ, it is not uncommon for the predictive performance of a model to deteriorate when the model is tested with patients other than those with whom it was developed. This has been recognized in the case of the PPS — due possibly to differences in patient cohort characteristics, location of care, and misunderstandings related to the use of the performance tool and the inter-reviewer discrepancy (18, 27). The differences between our cohort and the cohort used in the development of Prognostat are pronounced in terms age at treatment, cancer status, and PPS score.

However, we believe that instead of developing a new model, we should use knowledge from previous studies to update the existing prediction model by means of shrinkage and recalibration methods (28, 29). Updating methods can range from making adjustments to baseline survival to making adjustments to predictor weights using adjustment factors. This may entail re-estimating predictor weights and adding new predictors or removing existing predictors from the original model (10). Ideally, the updated model would also be externally validated. For Prognostat to be useful to

hospice and palliative care researchers, it should report explicit risk scores to be combined with new patient information and provide guidance on how this should be done.

Prognostat is also restricted in the framework of the Cox proportional hazards model, especially due to the fact that it is impossible to directly model and report the baseline survival function. This is essential in calibrating survival estimates for a new population of patients. We have found that the Royston-Parmar family of survival functions (30) is more accurate and flexible than the Cox proportional hazards model (31), as it allows for parametric modelling of the baseline survival function and relaxing of the proportional hazards assumption.

LIMITATION

A limitation of our study is that it was confined to external validation of an existing model, which needs to be recalibrated and tested prospectively on a data set independent from our patient population. Without explicit information from Prognostat regarding patient risk scores and linear predictors, this is not feasible at this time.

ACKNOWLEDGEMENTS

This study was supported in part by a United States Army Medical Research and Materiel Command grant under the program Development of Evidence-Based Clinical Decision Support System to Aid Prognostication in Terminally Ill Patients (DOA W81 XWH-09-0175).

Received: September 5, 2012
Final version accepted: December 12, 2013

REFERENCES

1. Steinhauser KE, Christakis NA, Clipp EC, et al. Factors considered important at the end of life by patients, family, physicians, and other care providers. JAMA 2000; 284(19): 2476-2482.
2. Glare P, Virik K, Jones M, et al. A systematic review of physicians' survival predictions in terminally ill cancer patients. BMJ 2003; 327(7408): 195-198.

3. Tanneberger S, Malavasi I, Mariano P, et al. Planning palliative or terminal care: the dilemma of doctors' prognoses in terminally ill cancer patients. Ann Oncol 2002; 13(8): 1320-1322.
4. Viganò A, Dorgan M, Bruera E, et al. The relative accuracy of the clinical estimation of the duration of life for patients with end of life cancer. Cancer 1999; 86(1): 170-176.
5. Christakis NA, Lamont EB. Extent and determinants of error in doctors' prognoses in terminally ill patients: prospective cohort study. BMJ 2000; 320(7233): 469-472.
6. Royston P, Moons S, Altman DG, et al. Prognosis and prognostic research: developing a prognostic model. BMJ 2009; 338: b604.
7. Altman DG, Royston P. What do we mean by validating a prognostic model? Stat Med 2000; 19(4): 453-473.
8. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010; 21(1): 128-138.
9. Altman DG, Vergouwe Y, Royston P, et al. Prognosis and prognostic research: validating a prognostic model. BMJ 2009; 338: b605.
10. Moons KG, Altman DG, Vergouwe Y, et al. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ 2009; 338: b606.
11. Vickers AJ. Prediction models: revolutionary in principle, but do they do more good than harm? J Clin Oncol 2011; 29(22): 2951-2952.
12. University of Victoria Health Terminology Group. Prognostat. Accessed August 28, 2012. http://htg.his.uvic.ca/tools/PrognosticTools/PalliativePerformanceScale/Prognostat/index.php
13. Anderson F, Downing GM, Hill J, et al. Palliative performance scale (PPS): a new tool. J Palliat Care 1996; 12(1): 5-11.
14. Fainsinger RL, Demoissac D, Cole J, et al. Home versus hospice inpatient care: discharge characteristics of palliative care patients in an acute care hospital. J Palliat Care 2000; 16(1): 29-34.
15. Morita T, Tsunoda J, Inoue S, et al. Effects of high dose opioids and sedatives on survival in terminally ill cancer patients. J Pain Symptom Manage 2001; 21(4): 282-289.
16. Virik K, Glare P. Validation of the Palliative Performance Scale for inpatients admitted to a palliative care unit in Sydney, Australia. J Pain Symptom Manage 2002; 23(6): 455-457.
17. Morita T, Tsunoda J, Inoue S, et al. Validity of the Palliative Performance Scale from a survival perspective. J Pain Symptom Manage 1999; 18(1): 2-3.
18. Lau F, Bell H, Dean M, et al. Use of the Palliative Performance Scale in survival prediction for terminally ill patients in Western Newfoundland, Canada. J Palliat Care 2008; 24(4): 282-284.
19. Lau F, Maida V, Downing M, et al. Use of the Palliative Performance Scale (PPS) for end-of-life prognostication in a palliative medicine consultation service. J Pain Symptom Manage 2009; 37(6): 965-972.
20. Lau F, Downing M, Lesperance M, et al. Using the Palliative Performance Scale to provide meaningful survival estimates. J Pain Symptom Manage 2009; 38(1): 134-144.
21. Harrold J, Rickerson E, Carroll JT, et al. Is the Palliative Performance Scale a useful predictor of mortality in a heterogeneous hospice population? J Palliat Med 2005; 8(3): 503-509.
22. Downing M, Lau F, Lesperance M, et al. Meta-analysis of survival prediction with Palliative Performance Scale. J Palliat Care 2007; 23(4): 245-252.
23. Graf E, Schmoor C, Sauerbrei W, et al. Assessment and comparison of prognostic classification schemes for survival data. Stat Med 1999; 18(17-18): 2529-2545.
24. Hu B, Palta M, Shao J. Properties of R(2) statistics for logistic regression. Stat Med 2006; 25(8): 1383-1395.
25. Hosmer DW, Lemeshow S, Sturdivant RX. Applied logistic regression. 2nd ed. New York: Wiley; 2000. (Wiley series in probability and statistics).

26. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982; 143(1): 29-36.
27. Lau F, Downing GM, Lesperance M, et al. Use of Palliative Performance Scale in end-of-life prognostication. J Palliat Med 2006; 9(5): 1066-1075.
28. Janssen KJ, Moons KG, Kalkman CJ, et al. Updating methods improved the performance of a clinical prediction model in new patients. J Clin Epidemiol 2008; 61(1): 76-86.
29. Toll DB, Janssen KJ, Vergouwe Y, et al. Validation, updating and impact of clinical prediction rules: a review. J Clin Epidemiol 2008; 61(11): 1085-1094.
30. Royston P, Parmar MK. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Stat Med 2002; 21(15): 2175-2197.
31. Miladinovic B, Kumar A, Mhaskar R, et al. A flexible alternative to the Cox proportional hazards model for assessing the prognostic accuracy of hospice patient survival. PLoS One 2012; 7(10): e47804.
32. Royston P, Parmar MK, Altman DG. External validation and updating of a prognostic survival model. Department of Statistical Science, University College London; 2010 Mar. 17. Accessed June 1, 2013. http://www.ucl.ac.uk/statistics/research/pdfs/rr307.pdf
33. Royston P. Flexible parametric survival analysis using Stata: beyond the Cox model. College Station (TX): Stata Press; 2011.
34. Royston P. Flexible parametric alternatives to the Cox model, and more. Stata J 2001; 1(1): 1-28.
35. Lambert PC, Royston P. Further development of flexible parametric models for survival analysis. Stata J 2009; 9(2): 265-290.
36. Jackson C. Flexible parametric survival models. Cambridge (UK): author; 2013 May 15. Accessed January 31, 2012. http://cran.r-project.org/web/packages/flexsurv/index.html

NOTE

[i] For a vector of covariates **x** and parameter vector **β**, the survival function S (t; **x**) for the Cox proportional hazards model is commonly expressed as $S(t;\mathbf{x}) = [S_0\ (t)]^{\exp(\mathbf{x\beta})}$ where S_0 (t) is the baseline survival function, i.e. survival function when all the covariates **x** are equal to zero. In the CPH framework, the estimation of the (linear) prognostic index **xβ** does not require the formulation of the baseline cumulative survival function S_0 (t), which itself can be estimated conditional on the covariate estimates using the Breslow and Kalbfleisch-Prentice estimators. However, the full parametric estimation of S_0 (t) is not possible, which makes prediction of baseline survival from the primary to the secondary data set not viable.

An alternative to CPH is the Royston-Parmar family of survival models, which relies on the transformation g(.), such that $g(S(t;x)) = g(S_0(t)) + x\beta$. The transformation g(.) can be either from the proportional hazard, proportional odds, Aranda-Ordaz or probit families. The baseline survival function $S_0(t)$ is approximated and smoothed by a restricted cubic spline function with m interior knots. A desirable feature of these functions is that unlike CPH, it can be reconstructed and used in post-validation model calibrating if the scale used (hazard, probit or odds), the knot positions, and the estimates of prognostic indices are reported. Calibration refers to estimating prognostic indices in the secondary data set using the parameter vector **β** estimated on the primary data set and applied to the vector of covariates **x** of the secondary data set. The interested reader is directed to a publication by Royston, Parmar, and Altman (32) for a detailed explanation. The methods can be implemented in Stata (33) statistical software using the *stpm* (34) and *stpm2* (35) commands, or in open source statistical software R as *flexsurv* package (36).